QC 6 B619
UNIVERSITY OF CALIFORNIA, SAN DIEGO E
3 1822 01282 8083

ARY
UNIVERSITY OF
CALIFORNIA
SAN DIEGO

LIBRARY
UNIVERSITY OF
CALIFORNIA
SAN DIEGO

THE UNIVERSITY LIBRARY
UNIVERSITY OF CALIFORNIA, SAN DIEGO

QC 6 B619
UNIVERSITY OF CALIFORNIA, SAN DIEGO E
3 1822 01282 8083

QC6
B619
SSE

SCIENCE AND ENGINEERING LIBRARY
University of California, San Diego

Please Note: This item is subject to RECALL after one week.

DATE DUE

APR 4 1991

[illegible]

SE 7 (Rev. 7/82)

UCSD Libr.

THE ORIGIN, NATURE, AND INFLUENCE OF RELATIVITY

A SERIES OF MATHEMATICAL TEXTS

EDITED BY

EARLE RAYMOND HEDRICK

B A Bernstein
1926

THE ORIGIN, NATURE, AND INFLUENCE OF RELATIVITY

LOWELL INSTITUTE LECTURES
LOWELL INSTITUTE — BOSTON
AND
LOS ANGELES LECTURES
UNIVERSITY OF CALIFORNIA — SOUTHERN BRANCH

BY
GEORGE DAVID BIRKHOFF
PROFESSOR OF MATHEMATICS IN HARVARD UNIVERSITY

New York
THE MACMILLAN COMPANY
1925

All rights reserved

PRINTED IN THE UNITED STATES OF AMERICA

COPYRIGHT, 1925,
BY THE MACMILLAN COMPANY.

Set up and electrotyped. Published December, 1925.

Norwood Press
J. S. Cushing Co. — Berwick & Smith Co.
Norwood, Mass., U.S.A.

PREFACE

In the fall of 1923 I gave six lectures on relativity at the Lowell Institute, and nearly a year later, eight lectures on the same subject at the Southern Branch of the University of California. The degree of general interest manifested, even in the two more technical lectures given at Los Angeles, seemed to indicate the desirability of revising, extending, and unifying this material in book form. My friend Professor Hedrick encouraged me to carry out such a project, and the book now lies before the reader as part of the Macmillan series of mathematical publications of which he is editor.

The trend of the times is toward the deeper unification of scientific thought. In particular it is beginning to be realized that there are a few basic ideas out of which arises the imposing superstructure of the physical sciences. Without doubt the theory of relativity has been the greatest single influence in this direction, for it has thrown all physical ideas, including the sacrosanct concepts of space and time, into the fervid crucible of creative speculation. The theory of Einstein constitutes a revolutionary advance, comparable with that due to Copernicus, and seems equally likely to affect the direction of mathematical, physical, and philosophical development.

If I have succeeded in portraying clearly the origin, nature, and influence of these basic ideas, and their modification in the theory of relativity, a main purpose of the book will have been fulfilled. I hope too that I have been able to contribute somewhat to the analysis of such fundamental starting points. Otherwise the book will not be a real factor in promoting unity and simplicity of thought about the physical world. Perhaps the most specific contribution made lies in the formulation of a definite theory of the structure of space and time, of matter and electricity, given in Chapters VI and VII. Here only are to be found the few mathematical equations elsewhere avoided. Even the non-mathematical reader should be able to proceed from the historical and logical analysis of the basic ideas given in the earlier chapters, and read these two more technical ones, so as to perceive more clearly how these ideas can be incorporated in the terse and exact language of mathematical symbolism. He will then be in the best possible position to follow the estimate of the philosophical influence of relativity contained in the concluding chapter, where I have merely stated my general position without any attempt at philosophical finesse.

In my opinion the range of material covered is not beyond the comprehension of any one possessing a modicum of scientific training. Consequently I venture to hope that the book will prove useful in a mathematical, physical, or philosophical course covering this field, as well as of interest to the general scientific reader. The field is indeed a most fascinating one, and I believe a course of this character must be given in all of our colleges and universities if we are to be fair to our students.

G. D. BIRKHOFF

CAMBRIDGE, MASS.
November, 1925

CONTENTS

CHAPTER PAGE

CHAPTER I. EUCLID, NEWTON, FARADAY, AND EINSTEIN 1

CHAPTER II. THE NATURE OF SPACE AND TIME 25

CHAPTER III. THE OLD AND NEW THEORIES OF GRAVITATION 50

CHAPTER IV. THE EXPERIMENTAL TESTS OF RELATIVITY 72

CHAPTER V. SOME RELATIVISTIC PARADOXES AND THEIR EXPLANATION . . . 89

CHAPTER VI. SOME GENERAL PRINCIPLES IN THE NEW PHYSICS 109

CHAPTER VII. THE STRUCTURE OF MATTER IN THE NEW PHYSICS . . . 139

CHAPTER VIII. THE PHILOSOPHICAL INFLUENCE OF RELATIVITY 166

INDEX 183

THE ORIGIN, NATURE, AND INFLUENCE OF RELATIVITY

CHAPTER I

EUCLID, NEWTON, FARADAY, AND EINSTEIN

A VERY simple and precise aspect of experience is that of the spatio-temporal relation between material objects. The underlying concepts of space and time have deeply influenced physical, astronomical, philosophical, and even religious thought. A broad understanding of them appears to be indispensable for an adequate appreciation of the universe in which we live. The historical development of these concepts, culminating in the theory of relativity of Einstein, throws much light upon their nature. Perhaps the names of Euclid, Newton, Faraday, and Einstein, more than any others, are representative of the various stages in the evolution which has taken place.

The significance of number, time, and space must have been felt in the earliest period when man's mental power was comparable to that which he possesses today. Elementary facts about unity and multiplicity, succession in time, size, and shape, are presented at every turn by experience. They have a curious inevitability and completeness, characteristic of mathematical truth, which was certainly perceived in this period. For example, the fact that there are two chief luminaries, the sun and the moon, that they function in alternation, and that they resemble each other in size and in shape must have stirred human interest even then. Doubtless the nascent elements of various fields of science were known to primitive man. The ability to handle material objects involved a perception of some principles of mechanics; experience in dealing with hot and cold bodies led to a familiarity with properties of heat; and the necessary requirement of taking account of day and night, and of the seasons, brought about a grasp of a little elementary astronomy.

The systematization of the process of counting was the first great step towards that understanding of and control over nature which is characteristic of our day. At first there was only the process of counting by means of the fingers, stones, or other objects. Then came the naming of the numbers, as one, two, three, etc., skill in reckoning with them, and their use in commercial pursuits, the measurement of land, and astronomy. Nearly two thousand years before our era the Egyptians were very competent in the use of number. The priesthood of that time had the leisure and interest for occupying itself with scientific thought. A mathematical papyrus written by a scribe Ahmes, entitled "Directions for obtaining the knowledge of all dark things," witnesses to their arithmetic proficiency.

In their surveying, artistic designing, and astronomy the Egyptians became adept in geometric construction and must have been familiar with many truths of geometry. The right triangle with sides of three, four, and five units was probably known to them and used in surveying. A floor in tiles with the pattern below indicated to the eye that the square on the longest side of a right triangle with two equal sides has an area precisely equal to the sum of the two areas of the squares on the equal sides. Moreover, practical urgency as well as merest curiosity led them to estimate with considerable accuracy the ratio of the distance around a circle to the distance across it. The inscribed hexagon showed at once that the ratio exceeded three. Since that day long ago geometric knowledge has advanced very far and this ratio has been fully determined: it is exactly fourfold as great as the number obtained when 1 is diminished by $\frac{1}{3}$, then increased by $\frac{1}{5}$, diminished by $\frac{1}{7}$, and so on *ad infinitum*.

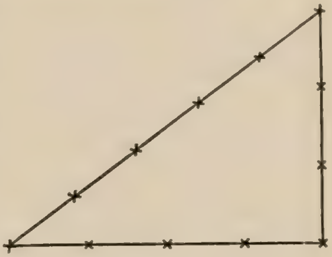
FIG. 1.

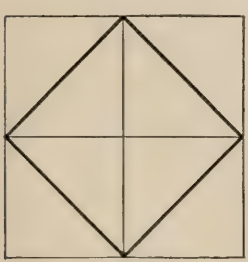
FIG. 2.

It may not be amiss to note in passing that not all elementary geometric conundrums have yet been answered by professional mathematicians. Thus, map-makers have noted that apparently any imaginable map on the plane or sphere can be colored in only four colors in such wise that two countries with a common boundary line have

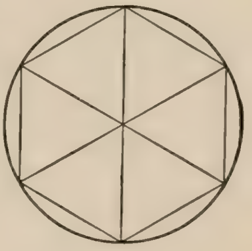
FIG. 3.

different colors. Despite persistent efforts, the truth of this conjecture has not yet been established, although five colors are known to be enough. Of like intriguing simplicity is the question raised a few years ago by the Japanese mathematician Kakeya as to the least area within which a line of given length can be turned around in a plane. An area only half as great as that of the circle with this length for diameter will suffice. No one has as yet been able to prove that this is the least possible area.

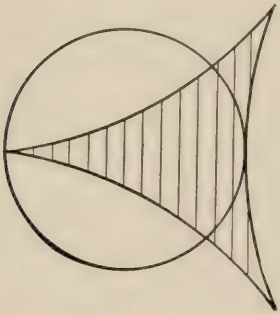

FIG. 4.

Since all physical science depends upon the foundation of mathematical truth, of which the discovery has been hastened by intellectual curiosity, a high place can be granted such curiosity for its proved value to mankind.

No advance in the logical analysis of geometric relations was achieved by the Egyptians. As soon as their empirical geometry had passed to Greece through the Greek geometer Thales about 600 B.C., it began to be transformed by Greek ability and interest in abstract speculation. The great although obscure figure of Pythagoras has unusual importance for the early Greek period. The relation between the sides of a right triangle, illustrated in the triangle of sides three, four, and five units, was probably discovered in its full generality by him. He found the two remaining regular solids of twelve and twenty sides; the Egyptians had known only the other three regular solids. To him is attributed the determination of the relation between the length of a musical string and its pitch, exemplified in a simple way by the fact that if its length is halved, the note is raised exactly an octave. For Pythagoras these truths indicated a deep and mystical interdependence between number and nature. Number was taken by him to be at the root of everything, and its properties formed the foundation of his philosophy. The regular solids represented elements in nature; the seven planets were held to give out musical notes proportionate to their distances from the sun, forming the music of the spheres. Thus the sway of quantitative law in the physical universe was discerned; the history of scientific thought affords no more inspiring example of prophetic insight.

Euclid, who lived about two centuries later, was the great systematizer of Greek geometry. His feeling for the purely logical side of geometry was uncanny in its accuracy, despite the fact that from the modern point of view his "*Elements*" are by no means free from blemishes. It was his achievement to have shown how all the known truths of geometry follow necessarily from a very small array of almost undeniable propositions called axioms and postulates. He was not primarily a discoverer of new geometric truths.

What interests us here is not so much the precise body of doctrine concerning points, straight lines, angles, circles, lengths, areas, etc., which he developed, as his point of view toward its application in the physical world.

In fact there are two alternative points of view towards geometric doctrine which can be held. One (that of Euclid) holds that we are dealing with a system of rigid bodies of fixed size and shape, of which material objects like tables, blackboards, rulers, the earth, etc., are fairly satisfactory embodiments. By handling these bodies and comparing them with one another, we arrive at the truths of geometry. For example, when it is said that two points determine one and only one straight line, this may be taken to mean that if we fix upon two points *A* and *B* of a rigid body, then when the simple kind of rigid body called a straight line, and exemplified by any ruler, is placed upon the rigid body so as to contain the points *A* and *B*, the line will always be found to contain the same points of the rigid body, forming the straight line *AB* of the body.

It is entirely compatible with this interpretation that any rigid body can be indefinitely extended into a "space." Thus a man in a moving railway train may estimate that a boulder directly opposite his window is 100 feet away and he may obtain its instantaneous position in the space attached to the train. The boulder also occupies a definite position in the space attached to the earth. Either of these spaces is in motion relative to the other.

The other point of view holds that there is a particular rigid body for which extension into a space is peculiarly proper, which space may then be called *absolute space* — a kind of omnipresent, penetrable, rigid body. Any material body fills a definite part of absolute space at any instant. With this interpretation, the fact stated about straight lines means only that for a given pair of distinct points *A* and *B* of this space there is one and only one set of its points which is of the type called a straight line and which contains these two points.

It was undoubtedly the second point of view that was generally held by the Greeks. Thus the point was a monad having position for the Pythagoreans. The conception of the point as that which is indivisible and has position was maintained also by Aristotle. Those who hold to the second point of view must contend with Aristotle that the principle of superposition has no place in geometry as a first principle, since two bodies may be superposed if and only if they occupy equal spaces.

On the contrary, Euclid himself, with that marvelous instinct for the essential which is characteristic of his work, defines the point as that which has no parts, *i.e.*, as the indivisible rigid body, and does not require that it have position in some hypothetical absolute space. This fact together with the use which he makes of superposition as a fundamental principle, indicates that he preferred not to employ the concept of absolute space.

It is necessary to have the divergence clearly in mind.

Euclid saw that the science of geometry as applied in nature attaches no particular importance to any special reference body. Thus for him a point was merely a small identifiable rigid body. Rigid bodies were to be compared by superposition, and it did not make any difference just how the superposition was effected in any particular case. Consequently, to insist that any particular body defined an absolute space was unnecessary. It seems, then, that Euclid adopted a relativistic point of view toward space.

On the other hand Aristotle saw that in everyday experience a particular reference body, namely, the earth, played a preponderating role. It must have appeared to him that other rigid bodies did not have comparable importance. A sufficient cause for this view was the generally accepted belief of his day that the air was a medium extending out from the earth indefinitely, and offering resistance to bodies moving in it. Despite the fact that from a purely logical standpoint all rigid bodies stand on a parity in the geometry of Euclid, some such vague considerations, not geometric but rather mechanical in their nature, inclined Aristotle to maintain the concept of an absolute space attached to the earth.

The opinion advanced here is that if some particular reference body has a superiority for the elucidation of nature, then it is desirable to signalize the fact by giving a special name to the space attached to it. It is undesirable to pay any veneration to that space, since the fact is then obscured that, from the standpoint of geometry in itself, all rigid bodies are on an equal footing.

There is, then, a relativity present in the application of the geometry of Euclid to nature which allows us to start from any particular rigid body of reference, as defining a space of its own, and then to locate the position of other material objects at any time relative to that space. The relativity found here may be called *geometric relativity*.

Before leaving the consideration of the Greek period, it is desirable to summarize the application of geometry in astronomical theory, for here the concept of time enters to join that of space, thus furnishing a very good vehicle for the statement and comprehension of the observed facts.

As might be expected, the transmission of light was thought by the ancients to be instantaneous. Nearly two thousand years elapsed before it was discovered that light travels at the rate of nearly 200,000 miles a second. Consequently, events were taken as happening when they were seen, this being done for events observed in the heavens as well as for events on the earth. Without any analysis of the notion of absolute time implicitly involved, a particular reference body and its absolute space were taken, to which the position of other bodies was then referred at any instant of absolute time. The explanation of astronomical phenomena was sought on the basis of this structure of space and time.

The application of geometry led very early to the recognition of the fact that the earth is a sphere. While traveling in Egypt Pythagoras became acquainted with this opinion. He seemed to have held secretly the heliocentric theory, which gave the sun a central place. However, the dominant opinion of the Greek period took the earth to be absolutely at rest, *i.e.*, held the earth to be the venerated reference body. The great observing astronomers of near the beginning of our era, Hipparchus and Ptolemy, chose this theory. Ptolemy held any other theory to be ridiculous, despite the fact that he recognized the superior simplicity of the heliocentric theory for some purposes. While the geocentric theory, in which the earth defined the absolute space, had the advantage that observations were made in that space, it had the disadvantage that the description of the motions of the sun, moon, and planets became very complicated. The concept of multiple epicyclic motion was invented to account for the facts in geocentric terms. To explain the motion of a planet on this basis, it was necessary to take account successively of rotations in three distinct spaces.

In the sixteenth century of our era, Copernicus selected the sun (taken as not rotating in the frame of the fixed stars) as defining the reference space. The theory of Copernicus found only slow acceptance. From our point of view, the real significance of his theory is that the explanation of the astronomical facts is in this way made very much more simple. No doubt the Copernican theory would have been accepted at once by Euclid, for none would have been more ready than he to grant that the new theory was accommodated to pure geometry fully as well as the Ptolemaic theory.

In the seventeenth century, Kepler and others, under the stimulating simplicity introduced by the Copernican ideas, were able to understand and to follow the motions of the moon and planets far more accurately than had been done before. There arose the belief that the same gravitational force which pulled terrestrial bodies to the earth was also operative between bodies at great distances apart, and controlled the solar system. However, the mathematical means for investigating the truth of this conjecture had not yet been invented.

The discovery of analytic geometry by Descartes in the first half of the seventeenth century was a first necessary step in this direction; the Pythagorean vision that number is at the root of everything was thereby justified in the field of geometry. It was the achievement of Newton to have invented the infinitesimal calculus, which deals with the rates of variation of varying quantities, and to have applied it to the problems of the solar system. A brilliant result was his law of universal gravitation.

The dogmatic adherence of Newton to the doctrines of absolute space and time is obvious from the first scholium of his "*Principia.*" A translation of a few characteristic passages will show how seriously he believed in these.

> I. Absolute, true, and mathematical time, of itself, and from its own nature flows equably without regard to anything external, and by another name is called duration; relative, apparent, and common time is some sensible and external (whether accurate or unequable) measure of duration by means of motion. . . .
>
> II. Absolute space, in its own nature, without regard to anything external, remains always similar and immovable. Relative space is some movable dimension or measure of the absolute space.
>
> III. Place is a part of space which a body takes up, and is according to the space, either absolute or relative.
>
> IV. Absolute motion is the translation of a body from one absolute place into another; and relative motion, the translation from one relative place into another.

Of course Newton endeavored to justify his unqualified adherence to these doctrines. In fairness to him it must be said that he succeeded partially in doing so in the following manner. Imagine a dumb-bell tossed into empty space at such a distance from any other bodies that gravitational forces need not be considered. On the basis of Newton's laws of motion, it follows that there will be a tendency of the weighted ends to separate if there is any rotation, whereas if there is no such rotation, there will be no such tendency. Thus, *absolute rotation* can be detected by means of the slight elongation of the bar joining the weights. Hence there are mechanical methods for ascertaining absolute rotation. Direct experiment indicates that such rotation is merely rotation with reference to the so-called fixed stars.

The case of motion in a straight line with unvarying velocity stands on an entirely different basis in Newton's theory. There is given no principle by which such uniform rectilinear motion of an undisturbed body in empty space can be ascertained, and it is inherent in his entire system of thought that this is so. Newton himself realized the situation more or less vaguely; for he says in the same scholium: "It is indeed a matter of great difficulty to discover, and effectively to distinguish, the true motions of particular bodies from the apparent; because the parts of that immovable space in which those motions are performed, do by no means come under the observation of our senses. Yet the thing is not altogether desperate. . . ."

For example, on the basis of Newton's theory, the sun is an undisturbed body, at least if the minor disturbing gravitational forces of the planets and of the other heavenly bodies are ignored. According to his theory, the sun may be either at rest in the absolute space, or describing a straight line in that space with constant velocity. His formulation of mechanical principles is such, however, that if the motion of all the heavenly bodies be computed on the hypothesis that the sun is at rest, *i.e.*, using the space attached to the sun, the result predicted will be exactly the same as if the space of any other star than the sun is taken to be at rest.

The advance of the Newtonian over the Euclidean point of view may be stated fairly as follows. Newton accepts the truths of geometry, according to which there are as many spaces as there are real or ideal rigid bodies, and yet he notes that some spaces are more suitable than others for describing natural phenomena, namely, the spaces attached to rigid bodies moving freely in empty space and in such wise as to be subject to no tendency of their parts to separate. It is only with respect to such a space that his first law of motion holds: "Every body perseveres in its state of rest, or of uniform motion in a straight line, unless it is compelled to change that state by forces impressed on it."

His doctrine of absolute space insists unnecessarily that one such space is supremely important, and accordingly to be termed absolute. But the fact is that all his laws of motion are expressed in identically the same way whatever such space is employed. An illustration of the truth of this fact is afforded by the behavior of a ball in the space of an elevator which is descending at a constant velocity. It will be found that the ball falls in this space exactly as though the elevator were at rest.

Newton's corpuscular theory of light and his law of gravitation also have this characteristic property of relativity.

If the space of a rotating body is selected, however, any freely moving body describes a complicated spiraliform curve in it, while the parts of the body itself tend to separate from one another. Thus the choice by Newton of reference spaces attached to a special type of freely moving body, namely, any one not in absolute rotation, is justified by a twofold test of simplicity. There is no justification for the selection of a special one of the possible reference bodies of this type as against all others.

Consequently, on the basis of the Newtonian theory a *mechanical relativity* holds sway which, though it is not so inclusive as the geometric relativity signalized above, does not distinguish between the spaces of various bodies that are in uniform translational motion with respect to one another.

The success of the heliocentric theory turns upon the fact that the sun is far more nearly an undisturbed body in such uniform motion than the earth, the moon, or the planets. But the space used cannot rotate with the sun and so is of the Copernican type; otherwise the stars would travel in circular orbits in the space, contrary to the first law of motion. The real motive for the use of the heliocentric system is now reduced to its proper basis, namely, that with this type of reference body the motions observed can be described with maximum simplicity.

Undoubtedly Aristotle had a vague feeling for the fact that when mechanical ideas are admitted there is less relativity allowable than when only purely geometric considerations are taken into account. This is especially apparent when he classifies motions as being according to nature and contrary to nature. However, he made no analysis of the reasons for his choice of the earth as a uniquely determined reference body, and indeed could not have been expected to do so.

The full validation of the concepts of absolute space and time seemed to be at hand when the role of electricity and magnetism in nature began to be discovered. These phenomena were the latest to claim the attention of physicists, so that even magnetism was scarcely mentioned by Newton.

The name of Faraday stands foremost among those who have investigated the facts concerned. To him is due the first complete statement of the interrelation between electrical and magnetic phenomena. He was the only one of the four men towards whom we are directing especial attention who was not a working mathematician. He lamented his lack greatly. It required his mathematical interpreter Maxwell to give Faraday's concretely expressed "lines of force" an exact and adequate statement. As Maxwell says of Faraday in the preface of "*Electricity and Magnetism*":

> "Faraday, in his mind's eye, saw lines of force traversing all space where the mathematicians saw centers of force attracting at a distance: Faraday saw a medium where they saw nothing but distance: Faraday sought the seat of the phenomena in real actions going on in the medium, they were satisfied that they had found it in a power of action at a distance impressed on the electric fluids."

When Maxwell had succeeded in giving the concepts of Faraday satisfactory mathematical formulation, he found that it followed also that the electromagnetic wave, such as is sent out from any wireless station today, travels of necessity with one and the same velocity under all circumstances in empty space, and that this velocity is given as the ratio of two well-known units in electricity and magnetism. Of course, the velocity is relative to the medium imagined by Faraday, called the ether and conceived as a particular one of the Newtonian spaces, namely, the one which Newton signalized as " absolute." Maxwell found that this calculated velocity was the same as the known velocity of light. Thence he derived the epoch-making conclusion that the light wave is electromagnetic, *i.e.*, the same in type as the wireless wave. All subsequent investigations have verified the truth of this conjecture and also the adequacy of his mathematical equations for expressing the facts of light and electromagnetism. Faraday had conjectured a relation between light and electricity. Maxwell demonstrated it in all detail.

At last the absolute space of the Newtonian philosophy, which had previously seemed without substance, began to have definite properties. At last there appeared to be a basis of distinction between the various mechanically equivalent spaces of Newton by means of the electromagnetic ether. Thus in turn the geometric concepts of Euclid, the mechanical concepts of Newton, and the concepts of electric and magnetic interaction of Faraday narrowed the possibilities down so that all spatial relativity was apparently excluded.

There were many who believed that the main laws of the physical world had been formulated, and that nothing remained but to find the precise way in which atomic matter and electricity were united in the atom.

As in any genuine theory, this view of nature was left open to verification. It became possible to determine absolute motion through the ether by experiment, and so discover absolute space. The fundamental method suggested for this ultimate quest can be explained as follows: If a man swims up-stream a given distance and back, he will find that it takes him more time to do so than if the water is at rest. This is particularly apparent in case he makes only slight headway against the stream. It is assumed that his velocity through the water is the same in both cases. Now a rigid body in motion through the ether may be compared to the banks of the stream, the ether to the stream which is moving past the body, and a beam of light which is flashed from a point of the rigid body to a second point at a given distance from the first and reflected back again, to the swimmer. It ought to take more time for the beam to travel back and forth along a bar in uniform motion in the ether than if the bar is at rest.

Owing to the variable character of the earth's motion, it was found possible to devise apparatus by which this increase of time should be revealed by a shift in position of certain interference fringes. As carefully tested in this way by Michelson and Morley in 1887, the earth showed not the slightest trace of motion relative to the ether. The attempt to explain the negative result by means of the hypothesis of a " drift " of the ether with the earth was meaningless, since the ether was by hypothesis a particular one of the Newtonian spaces, and the equations of Maxwell called for one and the same velocity of propagation of light in empty space of this sort under all circumstances. Such a suggestion of ether drift must remain incomprehensible until some modification of the foundations of physical theory is made. No one has proposed a definite and consistent modification of this kind.

A variety of subsequent experiments confirmed the general conclusion that the earth was entirely indifferent to motion relative to the ether. Nature unexpectedly refused to yield up the long-held secret of an absolute space at the very moment when all were of the opinion that it was to be wrested from her.

The perplexity into which physicists were thrown by the situation was much more embarrassing than appeared at first sight. Absolute space was intangible, and it began to appear that absolute time was not less so. If the synchronizing of clocks at different places on the earth's surface is to be effected accurately, a method of light-signaling or other electromagnetic signaling must naturally be employed. Without a knowledge of the velocity of the earth in the ether it was not possible to allow exactly for the time required by the signaling process. In fact the natural way of synchronizing such clocks appeared to be to assume the earth at rest in the ether, because of the absence of any effect; and this was out of the question inasmuch as then the timing would not be absolute but only relative to the earth.

The notion of absolute time has been taken without the slightest analysis in the historical sketch attempted above. It is necessary now to begin to consider ideas of time as well as of space, and hence it must be pointed out what a sweeping hypothesis is involved in the notion of absolute time. Suppose that I snap my fingers — an event. Other events happen throughout the universe which I propose to bundle together as having happened at the same instant. The method which I use to determine the events happening at the instant must be physically unique and not arbitrary in the least degree. If the physical means adopted of finding these events has any uncertainty in it, the measurement of distances will share in the uncertainty, since the position of a point of a body at one instant of time will be compared with the position of a second point at a different instant.

At first thought it might occur to me to correlate events as happening when I see them, for this is the method of ordinary experience. If I did that, however, I would consider a change in the light from a distant star to take place when it becomes visible through the telescope, contrary to all astronomical procedure.

For these reasons it is clear that the notion of absolute time involves a definite hypothesis concerning the physical structure of the universe.

A variety of physicists and mathematicians, among whom Fitzgerald, Lorentz, and Henri Poincaré are particularly to be mentioned, speculated upon the difficulty, without avoiding the unnecessary hypothesis.

In 1904 Lorentz arrived at a kind of solution of the problem, which it is most instructive to consider. He assumed that the circumstance of motion relative to the ether caused a moving body to contract in the direction of its motion by a part of its length (a few inches in the case of the earth), and a moving clock to slow down in the same ratio, precisely so as to account for the result of the Michelson-Morley experiment. It was as though the swimmer alluded to were required to go up-stream a shorter distance and the clock were running a little slowly, so that he would be considered to take exactly the same time in both cases. This explanation accounted for all the experimental results which offered difficulty, and seemed natural enough, because, if the ether is a real medium, it may be expected to affect the physical properties of measuring sticks and clocks moving in it.

The hypothesis of contraction presents two curious features. The first one is that it is incompatible with the ordinary concept of the rigid body as of fixed dimensions under all conditions. Henceforth only rigid bodies at rest are held to be of fixed dimensions.

The second feature is even more significant and deserves particular attention. It appears that if experiments are carried out on the body in uniform motion as if it were at rest, so that its length is held to be the same as when at rest, then the laws observed to hold are identically the same as if the body is truly at rest in absolute space. On account of the slowing down of clocks and the contraction of lengths as stated, it is at once clear that the velocity of light will appear to be the same for the reference system in motion as for the reference system at rest. The same result holds throughout.

This situation indicates nothing else than that it is experimentally impossible to distinguish between absolute rest and uniform motion in a straight line. Relativity, after being denied a position in the theory by general consent, is entering upon the stage without that consent. The relativity is, however, of a new type because of the fact that the time as well as the space of one reference body is distinct from that of another in relative motion. Of course any two such spatio-temporal systems of reference deal with the same totality of events. It is as if there were two systems of latitude and longitude on the earth for specifying position, for example, the usual system and a second in which the pole is slightly displaced. Both would deal with the same points on the earth. But latitude and longitude would be mixed up, just as space and time are in the two spatio-temporal systems of reference. In his theory Lorentz used the term *local time* to describe the time at the moving body.

If, despite the fact just referred to, according to which the various reference systems are physically indistinguishable, one deliberately prefers to single out a special reference system and call it " absolute," there can be no objection, any more than there can be to any arbitrary harmless act of the will. However, it is a privilege without physical significance.

Einstein was the first to perceive the true nature of the situation which was giving difficulty, and formulated the appropriate principle of relativity in 1905. The great stroke of critical insight on his part was to be able and willing to abandon the notion of absolute simultaneity which had never been questioned before. It has been stated that the notion of absolute simultaneity implies a significant correlation of events. If the velocity of light were infinite, the correlation would be immediate. It is finite, however, and the experimental facts indicate that the space and time of any undisturbed non-rotating body, *dealt with as though it were at rest*, have exactly the same physical properties as the space and time of another such body similarly dealt with. Why not take as equally valid all these spaces and times, and then attempt to correlate them without employing the hypothesis of absolute simultaneity? The new kind of relativity obtained, in which all such spaces and times appear on an equal footing, and not only mechanical properties of matter but also electromagnetic properties are taken into account, may be termed *optical relativity.* This relativity is spatio-temporal in character.

It is possible now to review the general course of the development of the concepts of space and time. At the outset events were taken to happen as seen, and the notion of absolute simultaneity was accepted as self-evident truth, which was natural in the circumstances. The first treatment of space turned on the Euclidean analysis of the rigid body and of the attached space which it defines. Any body was equally suitable as reference body, so that (for example) it was legitimate to use either the space of the earth or the space of the sun as the reference space. This was the geometric relativity of Euclid.

A more careful analysis of bodies, which took account of their mechanical properties, although using the concept of the space attached to a reference body at any instant, made it possible for Newton to go a step further and eliminate rotational relativity. Thus, the sun in the frame of the fixed stars was the superior body of reference in the solar system because it was nearly in the state of undisturbed motion without rotation, which considerations of convenience required of the reference body. Any other star than the sun would serve as well as the sun in yielding a space of reference of the best Newtonian type. Herein lay the mechanical relativity of Newton.

Then came the last step in advance, made possible only by the realization that, just because light travels at a finite velocity, absolute simultaneity is meaningless unless defined by a physical process. If the Newtonian relativity is slightly modified so as to admit the relativity of time, as well as of space, to any reference body in an undisturbed non-rotating state, a more homogeneous treatment of the physical universe is obtained, from the very nature of which it becomes impossible to choose any particular such space and time as possessing a superior validity. Such is the optical relativity of Einstein.

Thus, after space and absolute time have been used in the account of nature for nearly two thousand years, it has been found necessary to replace these concepts by a union of the two, called space-time. Under ordinary circumstances the old and new theories differ inappreciably in their predictions, and yet only the new theory will be found to explain certain phenomena in a satisfactory manner.

The one outstanding idea which must be grasped before the new theory of relativity can be properly understood is that the possibility of absolute simultaneity is not self-evident, but must refer to some significant physical correlation of events. In fact, two events are simultaneous if the two selected clocks, used where and when the events happen, give the same reading. Obviously this is the physical definition of simultaneity, and its meaning is not absolute unless clocks can be synchronized by unambiguous physical processes.

CHAPTER II

THE NATURE OF SPACE AND TIME

WHILE the historical survey of the development of the concepts of space and time is of essential value, only a direct logical analysis on the basis of the known experimental facts gives a satisfactory understanding of their real nature. Such an analysis indicates clearly why the space-time of relativity can be considered more nearly adequate than the classical space and time, and yet equally concrete and simple. As a preliminary step in this direction, the elementary facts of everyday experience will be stated independently of any particular theory. These are the same, whatever ultimate theory may be required to explain the physical universe down to the last ascertainable detail.

A fundamental characteristic of everyday experience is its arrangement in the order of time. Sensations are experienced in succession, and are recalled in the same succession as that in which they were experienced. There is also a flow in external nature nearly parallel to the flow of the stream of consciousness in the individual. Indeed, for the events which come within the range of immediate personal experience, sensations and the corresponding events are correlated in time without further question. It is true that occasionally a slight readjustment of this naïve interpretation is required, as when thunder follows lightning and it is realized that sound takes time to travel. Owing to the extraordinarily large velocity of light, it is never necessary to make an allowance of the same sort for light in ordinary experience.

It is very important to grasp the truth that a complete and exact parallelism between sensations in the individual consciousness and events occurring throughout the physical universe, implying absolute simultaneity, cannot be established in any simple way. For example, the attempt to correlate events as happening when seen was abandoned when the successive eclipses of Jupiter's satellites were observed to take place at smaller intervals of time when Jupiter and the earth were approaching one another than when receding. This anomalous result was accounted for by Römer in 1675 by means of the finite velocity of light. The complete failure of such a correlation of events is brought out even more clearly by a hypothetical experiment such as the following. Imagine that two observers *A* and *B* on two mountain tops, at a distance of one hundred miles from each other, send out flashes by means of revolving mirrors so that *A* sees *B*'s flash simultaneously with his own. Then *B* will see *A*'s flash about one thousandth of a second later than his own, and not simultaneously.

Thus it appears that the notion of absolute simultaneity of events requires physical definition by means of a technical process. Our preconception that there exists such a significant correlation of events is based on the illusion of everyday experience that events happen when seen.

With these reservations, the notion of measurable time on the earth, which we shall call *earth-time,* may be accepted as a working hypothesis. This time is measured with maximum precision by means of clocks at rest on the earth. Highly technical methods are used for synchronizing clocks far apart on the earth's surface.

In the stream of consciousness external objects called material bodies are identified and have a peculiar permanence. Among these are various nearly rigid bodies, possessing nearly constant size and shape. A piece of steel furnishes a good example. The comparison of such bodies by direct superposition leads to the concept of distance as determined by means of the ruler. In fact, take two simple rigid bodies of the very special type called straight edges. These can be exactly superposed in a variety of ways, and in particular so that any point of one coincides with a prescribed point of the other. Now let A and B be two points of the first straight edge, and let A' and B' be the two superposed points of the other. Place A' on B, and let C be the point of the first which falls at B'. Thus we can mark the set of equidistant points A, B, C, . . . on the first straight edge, which can now be used as a ruler for measuring distances. Likewise, it is possible to construct a protractor for measuring angles. The laws of Euclidean geometry embody the results of approximate measurement in most con-

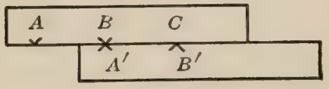

FIG. 5.

venient form. These laws are not to be taken as exactly applicable, particularly since no bodies are completely rigid.

An elaboration of these geometric ideas, in which the earth is taken as a rigid sphere of reference, gives rise to what may be called *earth-space.* The position of a body in earth-space may be specified by giving three spatial numbers, *e.g.*, latitude, longitude, and distance above sea level.

These systems of measurement of earth-space and earth-time may be usefully combined. Any terrestrial event may be completely specified by naming the instant of earth-time and the point of earth-space with which it is identified.

It is also evident from everyday experience that light is derived from definite material bodies like the sun, and occupies straight lines in earth-space in accordance with simple optical laws.

A physical theory will be held reasonable by the layman if it explains these properties of space, time, and light, as well as any further physical facts that may have come to his knowledge.

Our concern is not with such a superficial explanation of the obvious facts of everyday experience. It is entirely with that fundamental kind of explanation which aims to provide a complete and exact account not only of these facts but also of those discovered by the more searching processes of the laboratory and the observatory. How was it possible to advance in astronomy except by means of geometry which makes precise the spatial intuitions, and by use of the telescopic lens shaped with infinite care so as to supplement the feeble powers of direct observation? By such means the physical universe has been found to be subject to highly exact laws. As far as the underlying theory has been revealed to man, these laws appear to be unified and grandiose.

Let us turn then to examine spatio-temporal laws in some such spirit. It is well to begin here too by stating what are conceived to be the facts in the theory of absolute space and time hitherto accepted, and what are conceived to be the facts in the theory of relativity. In this way it will appear that there is a great deal in common to both, so that the divergences are minor by comparison and the new theory is equally simple and concrete. After so doing it will be possible to give intelligent consideration to the question of the logical analysis of space and time.

According to the classical theory there are two principal conditions which must be satisfied before the spatial comparison of material bodies (taken to be at the same temperature) becomes possible in any exact sense. All bodies in nature are elastic. Even a piece of steel lying upon a table is flattened by the force of gravity. Consequently, a first condition is that the bodies are small and at a great distance from any large bodies, so that such gravitational distortion is negligible. Furthermore, if such bodies are in a state of rotation, a similar condition of distortion will ensue because of the so-called centrifugal forces called into play. Hence it is necessary further that there be no rotation. Under these favorable circumstances any body may be said to be undisturbed. When two undisturbed bodies are placed in contact they will continue in contact, and exact measurement by superposition becomes possible. The laws of Euclidean geometry will then be found to hold with maximum exactitude, according to the classical theory.

It is precisely the same in the theory of relativity. Small bodies at a considerable distance from any large one may not be rotating, and then may be said to be in an undisturbed state. When placed in contact their measurement by superposition becomes possible, and the laws of Euclidean geometry will hold with the same degree of exactitude.

Even so, the degree of accuracy to be expected will not be complete if matter is atomic, because of atomic motion. It may be assumed that the atoms themselves are made up of perfectly elastic bodies carrying electric charges, and the difficulty will thereby be turned. However, our knowledge of the nature of the atom is so incomplete that this is not much more than a subterfuge. It must be admitted that neither the classical nor the relativistic theory gives any adequate account of the structure of matter.

Further measurement by means of light rays and Euclidean triangulation from any undisturbed reference body yields the position of other such bodies as if all these bodies were in an augmented body or medium in which the rays of light occupy the straight lines. Any point of this hypothetical medium may be called a point of space, whether or not an actual material body exists there. If there is no such body the space is said to be empty at the point. There is sometimes a tendency to find a lack of reality in the idea of empty space. This is due to a vague appreciation of the fact that space like number is an abstraction, although both are extremely useful. The physical properties noted when a material body is brought to a point of empty space are entirely definite. If a lighted match is placed in empty space (*i.e.*, in a vacuum), its flame is quenched; if a small balloon full of air is placed therein, it explodes.

It will be observed that the space which has been defined is the relative space about an undisturbed reference body. At this stage a difference of nomenclature between the classical and relativistic theories appears. Undisturbed bodies may be moving with respect to one another. In the relativistic theory the space defined by any one such body has precisely the same significance as any other, while in the classical theory a particular body is singled out as absolutely at rest, and is used to define an absolute space. Nevertheless, so far as the simple experimental methods used above are concerned, no difference will appear in the various spaces.

The laws of any such space have been stated to be Euclidean. It is distinctly worth while to pause and consider how simple and reasonable these laws are. The simplicity is often obscured by an effort at logical completeness, whereas if the logical demands imposed are only such as would have satisfied the early Greeks, geometric laws are very easily understood.

Any (undisturbed) body is made up of parts. The smallest such part is indivisible and is called a point. The simplest body containing two distinct points is a straight line, an example of which is afforded by any ruler. Two such straight lines have the characteristic property that they must coincide throughout if they coincide in two distinct points. The plane is the simplest body containing three distinct points not in a straight line. Two planes will coincide throughout if they coincide in three points not in a straight line.

Plane geometry deals wholly with the relations of the parts in a plane. The facts concerning geometry in the plane can be taken to repose upon the following four assumptions:

I. *Measurement of distance in a line can be made by means of the ruler.*

FIG. 6.

II. *Measurement of the angle between lines can be made by means of the protractor.*

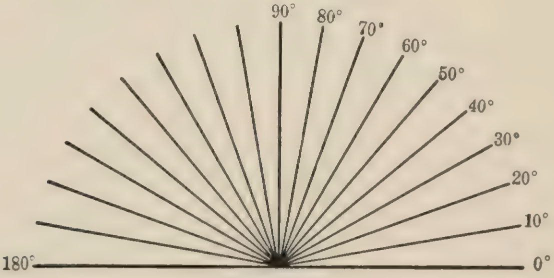

FIG. 7.

III. *One and only one straight line contains two given points.*

FIG. 8.

IV. *The plane is alike and even similar to itself in all its parts.*

It is only as applied to triangles to the extent indicated in the adjoining figure that this assumption is required.

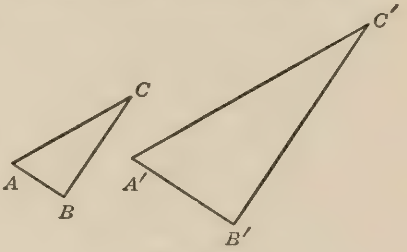

FIG. 9.

There are four propositions which follow readily from I–IV and are of central importance for the rest of geometry. We proceed to state and "prove" these briefly. Perhaps the first two of these will appear to be sufficiently evident without any proof.

V. *Two triangles with equal corresponding sides have equal corresponding angles.*

Consider first the special case of two such triangles ABC and ABD with the side AB in common, and with $AC = AD$, $BC = BD$, while the triangles lie upon opposite sides of AB. The two broken lines CAD, DAC, taken in the order named, have equal corresponding segments, and equal angles at A. Hence the triangles CAD and DAC have corresponding parts equal by IV. The angles labeled a in the figure are thus equal; likewise the angles labeled b are equal. It follows from II that the angles at C and D in the

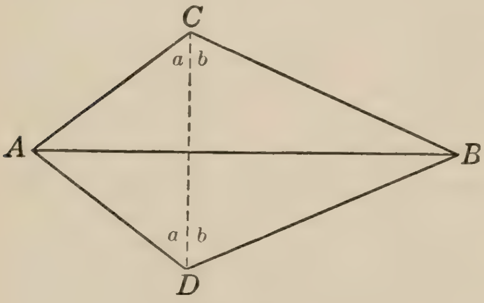

FIG. 10.

original triangles are equal, and from IV that these triangles must be equal in all respects.

Now suppose given any triangle $A'B'C'$ with sides equal to those of ABC. By use of IV lay off a triangle ABD equal in all respects to $A'B'C'$, and lying upon the opposite side of AB from ABC. By the above argument the triangle ABD, and thus $A'B'C'$, will be equal in all respects to the triangle ABC, as was to be proved.

VI. *The sum of the two angles less than* 90° *in a right triangle is* 90°.

The proof may be made as follows. Bisect the sides of the right triangle, by I. Join the three points of bisection by three straight lines, as may be done by III. The three small triangles touching the three corners of the given right triangle are then clearly similar to that triangle in consequence of IV, with corresponding sides half as great. The same is seen to be true of the interior small triangle by V. Hence the angles labeled a and b are the same, and it is apparent that when these are added by means of the protractor in accordance with II, their sum is 90°.

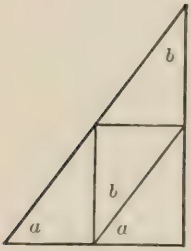

FIG. 11.

VII. *In a right triangle the side opposite the right angle is related always to the other two sides, as in the case of the* 3, 4, 5 *triangle when*

$$3^2 + 4^2 = 5^2, \qquad i.e., \qquad 9 + 16 = 25.$$

In the following discussion we shall show that a right triangle with sides adjacent to the right angle in the ratio 3 to 4, will have the side opposite that angle represented by 5. To prove it " in general " by the same method requires a little elementary algebra.

Consider the first triangle below, in which the sides adjacent to the right angle are represented by 9 and 12 and therefore in the ratio 3 to 4. The numbers 9 and 12 are chosen instead of 3 and 4 to avoid fractions later.

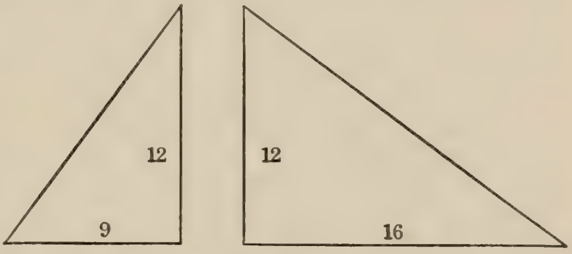

FIG. 12.

This triangle, in which the remaining side is to be found, is clearly similar to the second right triangle with sides adjacent to the right angle in the ratio 12 to 16, as follows from IV.

If these two triangles be joined together so that the common sides coincide, a larger triangle is formed, of which the largest side is 9 + 16 or 25 by I. The opposite angle is the sum of the two angles which appear in each of the pair of component triangles. Hence this opposite angle is 90° by VI, and the triangle formed by the combination is a right triangle also. The new right triangle thus obtained is similar to the previous ones, since its angles are the same. For if the smaller tri-

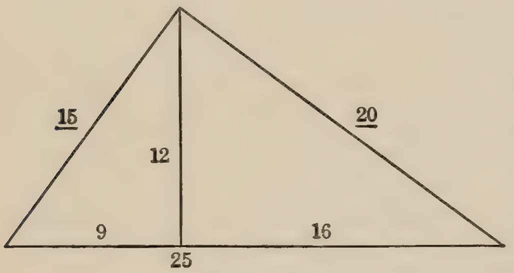

FIG. 13.

angles are magnified in the proper ratio by use of IV, and superposed on the large one by use of I, II, they will coincide with it throughout.

We are now sufficiently advanced in our knowledge to find the magnitude of the required third side of the first triangle. If that side is 15, the side 9 of the first triangle will be in the same ratio to this unknown side as that side, taken as belonging to the combined triangle, bears to the side 25 in that triangle, for $\frac{9}{15}$ and $\frac{15}{25}$ are both $\frac{3}{5}$. On the other hand, if that unknown side is more or less than 15, these ratios cannot be equal since one will be less than $\frac{3}{5}$ and the other greater. Thus the third side is exactly 15. The remaining side of the large triangle is of course 20. Hence it appears that if the sides adjacent to a right angle in a right triangle are represented by 3 and 4, the side opposite that angle is represented by 5.

The last proposition which we shall state is the following:

VIII. *Any two lines at right angles together with all the lines at right angles to them form a network of rectangles.*

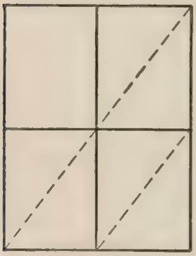

Fig. 14.

The truth of this statement is apparent from the same figure as we used in the proof of VI.

It will be found that with the aid of the propositions I–VIII it is a very easy matter to establish the main facts of elementary geometry which are really useful in its simplest applications.

The way in which many salient truths of geometry are incorporated in ordinary experience is astonishing. A host of relations between points, lines, planes, etc., are constantly apprehended at a single glance. Consider the obvious fact that if a man walks three miles east and then four miles north, he will arrive at the same destination as if he had first walked north four miles and then east three miles. This is a truth closely related to VIII. There are many similar obvious facts embodying geometric truths.

The preceding formulation of the methods and laws of spatial measurement indicates that there is no difference in the meaning of these for the old or the new theory. The same is true of temporal measurement in its most fundamental aspect, as will be explained immediately. It is only in the interrelation of space and time that a difference will be found to exist.

Let a tuning fork be in an undisturbed state in the space which it defines, and afterwards let it be set in a state of slight vibration. Each point of the tuning fork will then execute a periodic vibration about a point of that space, and thereby give a means of measurement of the lapse of time, *i.e.*, a *clock*. This is a perfect clock for any theory yet considered, so far as indicating lapse of time is concerned. A clock may be defined to be any small mechanism which serves to indicate equal intervals of time in its history. Clocks are omnipresent in nature, and the vibrating atom of any chemical element is a kind of ideal clock. A clock merely measures duration where it is, *i.e.*, local time.

The simplest method of comparing time at different places on an undisturbed reference body is by moving clocks very slowly from one place to another, so as not to interfere with their action. Still a second method is to read the clocks from a position halfway between them, in which case the readings should be the same. An alternative equivalent method available for a distant event is to take its time of happening as midway between the time of a light flash from the body reaching the point when the event happens, and the time of the reflected flash returning to the body. By these means simultaneity and time, as well as space, may be defined physically relative to the undisturbed reference body chosen.

The fact that events simultaneous for one reference body are not so for another can be made clear as follows. Imagine a long train in rapid motion on the earth, and that at a certain instant (in the space and time of the earth) a flash of light starts halfway along the train; this will be the case also for the flash in the space and time of the train. Obviously, by definition, the light flash will reach the ends of the train simultaneously in its space and time. But in the space and time of the earth the two events will not be simultaneous, for evidently the flash will reach the approaching end of the train before it reaches the receding end.

It is to be especially observed that the actual physical means employed in constructing the system of space and time reference are identical in both theories.

In nature no completely undisturbed bodies are to be found, although some are nearly in such a state. Some such body of reference is used to set up a specific system of space and time. The query naturally arises as to whether it is not possible to define space and time otherwise than by means of small undisturbed reference bodies. The answer is that the corrections to the preceding processes then required are extremely technical. Consider a heavy elastic sphere at rest in the absolute space of the classical theory; a first requisite for its use in space measurement is a complete knowledge of the theories of elasticity and gravitation. In all cases and no matter which theory is adopted, the path to highly exact spatio-temporal measurement otherwise than as stated is an extremely arduous one.

There is a slight difference in the spatio-temporal facts as these are conceived in the two theories. In the classical theory there is a particular undisturbed reference body, absolutely at rest, for which the statement of physical law assumes a superior simplicity. For it, light travels with the same velocity in all directions, and the shape and size of other undisturbed bodies in relative motion, taken at an instant of its own absolute time, are the same as if the bodies were at rest. In the theory of relativity, every undisturbed reference body defines a relative space and time which has the same validity as that of any other body. In particular the velocity of light is the same in all such spaces. To counterbalance this increase in simplicity, however, there will exist a modification in the shape and size of an undisturbed reference body in relative motion as obtained by the indirect optical method, although there will be no modification if the dimensions are determined by direct spatial measurement.

It is particularly important to understand the modification to which the notions of time and simultaneity are subjected in the passage from the old to the new ideas.

In whatever way time may be determined, there is one property unconditionally demanded, namely, that it be impossible to modify events subsequent to their happening. If a wireless signal is sent from Boston at exactly 12 o'clock, it must not affect the New York receiving station before that hour. Similarly, if a signal is sent from New York at 12 o'clock, it must not reach Boston before that hour. However, since the signal only takes about one five-hundredth of a second to pass back and forth, it is clear that this single demand upon the notion of time has fixed simultaneity to within very narrow limits. Now there is no known velocity which exceeds that of light and electricity. In fact, such a velocity seems contrary to the essential constitution of nature. Hence this kind of simultaneity appears to admit of no clearer definition. It may be termed ordinary simultaneity.

Absolute simultaneity is of an entirely different character. It implies the existence of a uniquely defined system of time measurement, with the aid of which the statement of general physical law obtains its maximum simplicity. For example, the law of gravitation formulated by Newton for classical physics will only hold if absolute time is used in the measurement of the gravitational acceleration of bodies. Such is the absolute simultaneity of classical physics.

On the other hand, if some specific system of time measurement is taken which yields the maximum simplicity in the statement of physical laws, and yet there exist other systems for which the same laws hold, then there is no such thing as absolute time or absolute simultaneity. This is the situation in relativity where the systems of time measurement for various undisturbed reference bodies are equally valid.

Thus the absolute time and simultaneity of the classical physics are replaced by a relative time and simultaneity dependent on the undisturbed reference body selected. The new simultaneity meets the fundamental demand first imposed.

There is a third kind of simultaneity which may be called transcendental. If the mind of Deity envisaged events in a temporal order, there would be defined a transcendental type of simultaneity. Similarly, if telepathic communication became possible and it was found to be instantaneous in the sense that immediate communication back and forth were possible, then a species of transcendental simultaneity would be thereby defined. The validity of any transcendental simultaneity is entirely speculative, of course.

Einstein was the first to realize that it is not necessary to hold to the notion of absolute time. While absolute space had not been securely defined in the Newtonian theory, absolute time and its attendant concept of absolute simultaneity had never been questioned before. The cornerstone of Einstein's logical analysis of space and time declares the abandonment of the unnecessary hypothesis of absolute time, at least until it is called for by the physical facts. This implies that the notion of a rigid body must be held in abeyance, too, since the rigid body is conceived of as one which can move about freely, while the distance between a pair of its points at any instant of absolute time remains invariable. Such a statement becomes meaningless when the notion of absolute time is abandoned.

It may ultimately appear that one particular reference space possesses an entirely superior character for the statement of the laws of nature, and hence we may wish to call that space absolute. Likewise some one time may be preferable, in which case it may be called absolute time. Nevertheless, even if this happens to be the case, it will be well not to forget that such a particular choice is in no sense *a priori*, but is a matter of convenience and is arrived at through observed laws experimentally determined. It will be seen that this broader point of view insists that the universe of events has primary significance, and that it is desirable to characterize it in some fashion or other in terms of a three-dimensional space and one-dimensional time, together constituting what may be called the four-dimensional space-time of events.

It has been made clear earlier that relativity has no useful application in the limited field of direct human observation. Here the ideas of earth-time and earth-space were found to provide the natural means of accounting for the directly observed facts. The application of relativity was found instead to lie in the complete and exact statement of physical laws. However, the very terms "earth-time" and "earth-space" indicate that space and time are relative. Hence there ought to be no difficulty in passing to the relativistic point of view. The detailed comparison just made indicates that the new theory is simpler in some ways than the old, and that both have a great deal in common.

We now propose, in conclusion of this chapter, to attempt to consider space and time without prejudice by looking afresh at the physical world about us. By doing so we shall find it possible to adopt a more affirmative point of view towards relativity.

It is idealized interstellar space which furnishes the *motif* for the new physics much as the idealized rigid body did for the old. In interstellar space are found vast stretches of empty space with here and there comparatively infinitesimal stars, moving with velocities relative to one another which may reach hundreds of miles a second. From these bodies light spreads out with inconceivable exactitude in every direction, and may be measured by means of that eye of the astronomer, the telescope, and his clock. Thus we are striving to interpret anew this world of the astronomer.

In the new model there seems to be only one type of measurable quantity, namely, that of duration at any one of the particles by which we propose to replace the stars and other bodies in interstellar space, in order to idealize and simplify it to the utmost. We propose to assume also that this duration is measurable by means of clocks at the particles, and has physical significance in the correlation of events happening at any particle. These clocks are supposed to be identical in structure, so that if they are brought together they will run at the same rate. Such clocks exist in any physical theory which has been so far proposed.

Other particles may be at rest relatively to a given particle, as will be evidenced when a light signal always takes the same time to travel back and forth in the time measured on the clock of the first particle. Three particles relatively at rest will furnish a basis triangle, by which a space of reference may be set up through a light-signaling process only. The experimental rules adopted in doing this will be exactly the same as for three particles at rest in the ether of the classical theory, so that a particular time of reference will be obtained also.

It can now be *proved* that if space-time is alike in all its parts (which seems as natural and inevitable an assumption as that the plane of Euclidean geometry is alike in all its parts) then the reference space and time will have notably simple properties.

In the first place, any such space will be subject to exact Euclidean laws of distances and of angles obtained by light-signaling methods. Here we obtain an optical interpretation of space akin to that demanded by Aristotle. Bodies at each instant of the chosen time fill a part of this space. It is found, further, that light will travel with one constant velocity in all directions, and other particles will travel with uniform velocity in a straight line, the velocity being less than that of light. It will even appear that the relative velocity of two particles is the same in the system of reference of either.

All the assumptions hitherto used are granted either in the classical or in the relativistic theory, and so it would seem that there is no denying the correctness of the logical analysis of the model presented by interstellar space, as far as we have proceeded with it.

The characteristic hypothesis of the theory of relativity now enters. Up to this point there has been a perfect parity between all the particles in interstellar space, and as far as that model is concerned it seems as inevitable to assume that this parity is complete as it does to assume that all points are on a parity in the Euclidean plane. The hypothesis of relativity merely asserts the parity to be complete.

It is desirable to be more specific. If a particle *B* is moving relatively to the reference particle *A*, it is not hard to show that the apparent rate of *B*'s clock measured in *A*'s own reference space and time will vary in general with *B*'s apparent velocity. It is necessary to emphasize the fact that the rate of *B*'s clock so measured cannot be in general the same as the rate of the clock at *A* even in the classical theory, since in that theory it will only be the same if the reference particle *A* happens to be absolutely at rest.

The characteristic assumption may now be stated in more specific terms. The apparent rate of *B*'s clock in *A*'s space and time must depend in exactly the same way on the relative velocity of *B*, whatever particle *A* is.

It is even sufficient to demand that when direct visual observation of the rate of *B*'s clock is made at *A*, its rate is the same as that of *A*'s clock as observed visually at *B*, at least if *A* and *B* approach one another or recede in the same straight line.

When stated in this form, the truth of relativity almost becomes a philosophic necessity. Imagine an interstellar space containing only two particles which are approaching each other. Why should not the relation between them be entirely symmetric?

Now the space-time was highly idealized in that only point particles were assumed to be present. As a matter of fact such matter will always have spatial extension. If the spirit of the characteristic hypothesis of relativity is to be maintained, it must be assumed that undisturbed elastic matter at rest in any reference space will always have the same size and shape, *i.e.*, that the back and forth signaling time between any pair of points of such a body will be the same under all circumstances. It follows that direct measurement by superposition of undisturbed elastic bodies may be employed. This substantiates more fully the claim made earlier concerning the role of geometry in relativity. Of course in actual experience the conditions are not met exactly, but the same difficulty is presented whatever theory is adopted.

A more careful examination of the consequences of the hypothesis of relativity shows that the respective times and spaces of any two reference bodies are the same if and only if they are relatively at rest. If they are moving slowly relatively to one another in comparison with the velocity of light, the differences between their respective spaces and times are extremely minute, and altogether beyond the range of observation under the conditions which usually obtain. Nevertheless, differences are found to exist, which justifies the caution which enjoined us from assuming an absolute time. It must be emphasized that the various spaces and times are effectively interrelated, so that time can no longer be thought of by itself.

It is unreasonable to expect any very obvious relation to exist between two different systems of space and time reference. Is it obvious how latitude and longitude on the earth would be modified if the magnetic pole were used instead of the north pole?

Thus the facts observed in everyday experience are as well accounted for by the theory of relativity as by the classical theory of space and time, while the former is superior in that it explains the negative result to be expected from any experiment designed to determine the absolute motion of a particle in interstellar space.

The theory outlined above is the special theory of relativity formulated by Einstein in 1905. The formulation took no account of gravitation, and yet it is possible to formulate a relativistic law of gravitation in harmony with that theory. Einstein's general theory of relativity of 1915, to which we turn in the next chapter, accounted for gravitation even more completely and profoundly.

It is often thought that there is something inconsistent about the special theory of Einstein. This is a mistaken point of view. In fact the theory of relativity, like Euclidean geometry, is a complete theory which may or may not be exactly applicable to the physical world in which we live. It is a self-consistent abstraction in any case, and the proof of its self-consistency is no more difficult than that of elementary geometry. The self-consistency of geometry is usually assumed without question because geometric relations are realized approximately in experience, but that circumstance alone is not sufficient to make it rigorously self-consistent. Similarly, if we are willing to take the analysis of interstellar space (which certainly has physical existence) as proceeding intuitionally along the lines indicated above, we will not require a proof of self-consistency. If we refuse to accept this approach, an argument may be made of the purely arithmetic type available in geometry.

In our own world in which the relative velocities of bodies are very small when compared with that of light, nearly all the spaces and times are substantially identical. It may be conjectured that the laws of nature are such that, although relative velocities were high in the remote past, they have tended to diminish. Thus the low relative velocities now found may be accounted for. It is natural, nevertheless, to look upon the situation just alluded to as an argument against relativity from the very standpoint of convenience so much insisted on hitherto, for it is certainly convenient to define some particular space and time with respect to which all the stars are in slow motion. The reply to this argument is merely that it is in no way indicated exactly what space and time are to be taken as absolute.

Thus the new point of view takes its start in the grandiose interstellar space of the astronomer, in which particles are in motion relative to one another, and light waves pass to and fro. The local clock and telescope are the only instruments of precision. If the parity of these particles be granted as a most reasonable hypothesis, the relativity of both space and time to the particular particle flows with the same inexorable necessity as prevails in Euclidean geometry.

The situation is akin to that presented by the lines of symmetry in a closed oval. If the oval is circular, there are infinitely many lines of symmetry passing through the center of the circle, and each is as valid as any other. The relativity of nature resides in a similar symmetry of natural law for various systems of spatio-temporal measurement. If a point on the circle is marked, there will be only one line of symmetry which will pass through the marked point. This is analogous to the fact that a specification of the reference body fixes the space and time completely.

CHAPTER III

THE OLD AND NEW THEORIES OF GRAVITATION

With these facts in mind concerning Einstein's special theory of relativity of 1905, it is possible for us to approach with advantage his general theory of 1915, which is a fundamental extension of the special theory, devised to account for the phenomenon of gravitation.

It is a commonplace of experience that when material objects are left free, they fall to the earth. If suspended from above they stretch downward, and if supported from below they flatten in the same direction. It is customary to think of such effects as due to gravitational force. This force acts with nearly the same intensity everywhere on the earth's surface. Very careful experiments indicate that it is proportional to the mass or quantity of matter concerned.

It may have been felt in early times that some force of an analogous nature keeps the heavenly bodies in their prescribed paths. Kepler had specific ideas about gravitation, nearly all of which were correct. He conjectured that the *virtus motrix* diminished as the distance between the bodies increased, in particular inversely as the distance. However, the mathematical tools necessary to deal with the questions at issue had not then been invented; these were the analytic geometry of Descartes and the infinitesimal calculus of Newton and Leibniz.

The possibility that the force varied inversely as the square of the distance, so that when the distance was doubled the force diminished to one quarter of its magnitude, interested Newton as well as several of his contemporaries. It was the extraordinary accomplishment of Newton to have established with the aid of the calculus that this law of force holds with a high degree of accuracy throughout nature, by applying it to the solar system, comets, tides, etc. The subsequent more thoroughgoing application to nature has required notable mathematical skill as well as extensive computation, and the agreement of theory and the facts has been found to be nearly complete.

The Newtonian law is unparalleled in the realm of physics because it calls for the instantaneous spreading of a disturbance throughout all space. Thus, if a material object is moved suddenly, the disturbance in gravitational force caused thereby is supposed to affect instantly the most distant stars. No explanation of the law has been forthcoming which accounts for it as a secondary result of the properties of matter and electricity. Substantially all that has been done is to state the law and deduce mathematically its many consequences.

It would be unreasonable to expect a complete explanation of a phenomenon as basic as gravitation, and yet there is much to be said that is well worthy of careful consideration. A complete explanation of any set of facts undertakes to account for them in terms of still more fundamental accepted facts. For instance, heat may be explained in this way as molecular motion, discerned in a gross statistical way by means of the senses. To whatever extent this kind of explanation may be carried, there must remain other facts, taken for granted and not explicable in the same satisfying manner. There is, however, another imperfect kind of explanation (such as is here attempted for gravitation) which may be possible. For instance, geometry can be developed from a few intuitively accepted truths into all its infinite and beautiful variety by means of systematic reasoning. Such a development gives an excellent example of the second type of explanation. Another illustration of the same type is found in the so-called analytic theory of heat, where a few reasonable assumptions, such as that heat is a measurable quantity and flows from hot to cold, yield a complete theory by logical development.

It will appear that the law of gravitation is similarly predetermined by its broad qualitative properties as an interaction between bodies not in direct contact, together with the principle of relativity present in the particular framework of the physical universe which is taken as starting point. This is true of the theories of gravitation of both Newton and Einstein. Consequently it would seem that the law of gravitation expresses in the simplest possible terms the direct purpose of nature that material bodies tend to approach one another in empty space.

In the framework of space and absolute time at the base of Newton's theories, any undisturbed body yields an attached reference space in which other undisturbed bodies move with uniform velocities in a straight line. Let us select some such space of reference. Any other space which moves relatively to the selected space with uniform rectilinear motion will serve equally well.

The type of relativity which is present in the special model forming the starting point is particularly to be noted. Here the bodies are so small and so far apart that they may be treated as point particles which do not disturb one another. The special model may be considered to be realized with high exactitude in interstellar space. Now what is the type of relativity referred to? It is that a particular space can be replaced by any other space moving uniformly in any direction, and also that particular units of length and time can be replaced by any other units, without change in the statement of physical laws. It is natural to try to maintain such a relativity in the generalization of the special model to space and time taken as the container of interacting matter. In another form, this requirement means that the physical properties of bodies are to be thought of as independent of velocity and orientation in the selected space.

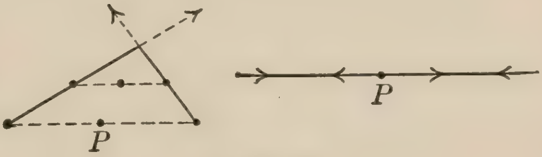

Fig. 15.

First let us consider briefly the contact forces such as arise when equal, rigid, spherical particles happen to collide. The two particles will move with constant velocities along straight lines before collision. The point P midway between their centers will do likewise. Let the space of reference be taken to be attached to this point, P. Then in the equally valid new reference space, the two particles will approach one another with equal velocities in opposite directions and collide at P. It is assumed for the sake of simplicity that the collision is direct and not glancing. The recoil must be in the straight line joining the two particles, if the principle of relativity which obtains in the special model is not to be destroyed. In fact, otherwise, the result would not be independent of orientation in the space selected. Likewise, the two velocities of recoil must be equal, or there would be a physical distinction between the two directions on the line of approach. Finally, if we interpret the relativity of time in the mathematical sense that the unit of time can be changed to its negative (*i.e.*, that any series of happenings can be described in the reverse order), it appears readily enough that the velocity of recoil must be the same as that of approach.

Thus the entire behavior under direct collision is determined, and, by reference back to the original space, it is found that the laws of collision are such as to leave unaltered the momentum and energy as well as the mass of the pair of colliding particles. By such means, with the natural and inevitable extension to gases, liquids, and solids, as obtained by aggregation of molecular particles, the behavior of bodies in contact can be determined in large measure by the mere requirement that the relativity of space and time remains as extensive as in the special model.

The alternative kind of force is that which acts between bodies through empty space. Let us see how it is that such gravitational force is determined in character by the same requirement that the relativity present in the special model be preserved.

It should be observed that the word *force* as used here and earlier has no metaphysical significance. When any particle departs from motion with constant velocity in a straight line, then, in the momentarily attached Newtonian space, the particle will be leaving its position by a definite distance in one second. Such a displacement in position for each unit of mass is the measure of the force which acts. Hence we can define force by purely geometric methods based on a measurement of displacement from uniform motion in a straight line, and this is the definition adopted here.

Now the Newtonian analysis of space and time is incomplete to the extent that it proceeds as if light were propagated instantaneously. But such a restriction only means that the velocities of bodies are of very small magnitude in comparison with the velocity of light. In order that such a situation be maintained, it is necessary that only bodies of small mass be considered. It is, therefore, a restriction to the case of small masses and velocities.

Take first two equal spheres of small mass relatively at rest, held at a fixed distance apart by a very light rigid bar. If the bar is removed, the mutual influence hypothesized must necessarily result in a displacement of the spheres along their line of centers, measurable as a force. This conclusion follows by the principle of relativity used before. The magnitude of the force can only depend on the distance between the spheres. Unless some particular absolute unit of length exists, the force must vary as a power of the distance, as a little mathematical argument would show; the existence of such a unit of length is inconsistent with the same principle of relativity. The force cannot vary directly as a power if it is to vanish when the two bodies are at a great distance from one another. It cannot vary inversely as the first power of the distance, for in that case bodies approaching one another from a distance acquire indefinitely large velocities by mutual influence. Nor can it vary inversely as the third power or a higher power, for then the mere aggregation of continuous matter provides an infinite supply of energy. Thus, by very obvious requirements, the range of choice appears effectively narrowed down to the law of Newton, according to which the force varies inversely as the second power of the distance.

Although only the gravitational interaction of two small equal spheres has been considered, the principle of the superposition of small effects, widely observed in nature, is of service in the determination of the forces of interaction between any number of equal or unequal small bodies. A simple illustration of this principle is afforded when sets of circular ripples, caused by as many pebbles striking the surface of the water at different points, meet. The height of the combined wave is the sum of the heights of the separate waves. The principle of superposition is one that holds by virtue of an inner mathematical necessity rather than an independent physical principle.

Hence the mechanical and gravitational laws of interaction are determined, once the Newtonian framework of space and absolute time is accepted, and the full relativity of that framework is insisted upon. The logical incompleteness in the preceding development of this fact is by no means to be ignored, but the fact itself seems undeniable.

One limitation in that theory arises from the circumstance that it takes no adequate account of the finiteness of the velocity of light. Accordingly, any inaccuracy in the law of gravitation obtained is likely to manifest itself when the relative velocities are high. In particular, it would be expected that, of all the planets, Mercury would deviate most from its predicted orbit, for it moves with the highest velocity of any planet under the solar attraction. Such a deviation has been observed.

Furthermore, when this framework is adopted, it is not possible to maintain the same degree of spatial relativity in dealing with electricity and magnetism. This is also to be anticipated, since light is electromagnetic, and the disregard of its velocity is particularly improper from the electromagnetic point of view. It thus becomes necessary to adopt the absolute space or ether of Faraday and Maxwell.

The diagrammatic circle of relativity may be used to illustrate the considerations adduced above. Consider the circle with the infinitely many symmetric reference lines through its center, representing the relativity which is present at the outset in the Newtonian framework. Suppose that a point in the plane of the circle is to be specified in such wise that the symmetry is not lost. The point must necessarily be the center. Or suppose that another circle is to be specified under a similar condition. It must have the same center as the first circle. On the other hand, if it is demanded that a line be laid down, the symmetry is necessarily destroyed. The mechanical and gravitational laws are analogous to the point and circle to be determined. The fact that these laws are in accord with the principle of relativity indicates the basic importance of that principle. On the other hand, the laws of electromagnetism are inconsistent with the principle, and so are akin to the straight line.

If the preceding analysis is correct, the heuristic value of the principle of relativity for classical physics can scarcely be overestimated. With it the form of the laws of mechanics and gravitation seem predetermined while without it these appear to be given at random. The laws of electricity and magnetism are not so predetermined, however, in that theory.

In passing to the special theory of relativity invented by Einstein in 1905, we begin with a more adequate analysis of space, time, and light, and obtain a corresponding framework. Any undisturbed particle defines a particular space and time, relative to itself, in which interacting matter is taken to be present. When it is required that the new, spatio-temporal type of relativity of the special model be strictly preserved, a very interesting situation results. The laws of contact action of elastic bodies can be determined by means like those available in the Newtonian case. Slightly modified principles of conservation of mass and energy, and momentum are obtained, in such wise that the theory agrees with the known facts of mechanics at the velocities found in nature.

The account given of the electromagnetic theory in empty space becomes extraordinarily symmetric with this type of relativity. The theory may be approached very readily by means of a few broad qualitative assumptions and by use of the principle of spatio-temporal relativity. The symmetry between space and time is so complete that one is justified in writing down the correct dimensional equation

$$186{,}300 \text{ miles} = \sqrt{-1} \text{ seconds}.$$

The meaning of such a mystical equation cannot be elaborated here: it indicates a formal symmetry between the properties of space and of time. This peculiar kind of symmetry is found throughout the theory.

If now we proceed to consider the interaction of bodies at a distance, it appears that, in the case of sparsely distributed matter at rest in some space, the law must be of the Newtonian type for the very reasons advanced before. An appropriate extension to the case of matter in relative motion may be obtained by the so-called principle of the permanence of mathematical form. In applying this principle, the simplest law is sought which is consistent with the principle of relativity and which reduces to that of Newton for such a static distribution of matter.

Any beginner in algebra who knows that a^1 stands for a, a^2 for a times a, a^3 for a^2 times a, etc., sees that a law of combination by addition of exponents holds, so that a^2 times a^3 equals a^5, for instance. If asked how to interpret $a^{\frac{1}{2}}$, he observes that by the same law of combination $a^{\frac{1}{2}}$ times $a^{\frac{1}{2}}$ ought to be a^1 or a. Hence he infers by the principle of permanence of mathematical form that $a^{\frac{1}{2}}$ ought to stand for the square root of a. This principle has been found of the greatest value for heuristic purposes, and can be extended remarkably.

It is an easy matter to set down the complete equations involved, so that even the non-mathematical reader who does not know the simple language of the calculus can see the high formal parallelism between the classical theory and that here proposed. To achieve simplicity, only free matter will be considered to be present.

In the classical theory there is given first and foremost the so-called potential equation,

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} + \frac{\partial^2 g}{\partial z^2} = -4\,\pi\rho$$

Here x, y, z are the space coordinates of a particle, ρ is the density of matter, and g is the gravitational potential function. This equation is consistent of course with the mechanical relativity of Newton. In fact the linear differential operator which appears on the left-hand side of the equation is the only one possessing this property.

In the special theory of relativity this is naturally replaced by

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} + \frac{\partial^2 g}{\partial z^2} - \frac{\partial^2 g}{\partial t^2} = -4\,\pi\rho,$$

t denoting the time, since the linear differential operator of the second order which now appears on the left is the only one consistent with the optical relativity of Einstein.

There are two further conditions in the classical theory, one expressing the conservation of mass, and the other that the acceleration of every particle is in the direction of the gravitational force. These are respectively

$$\frac{\partial \rho}{\partial t} + \rho \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) = 0,$$

and

$$\frac{\partial u}{\partial t} = \frac{\partial g}{\partial x}, \qquad \frac{\partial v}{\partial t} = \frac{\partial g}{\partial y}, \qquad \frac{\partial w}{\partial t} = \frac{\partial g}{\partial z},$$

provided that t, x, y, z are the coordinates relative to an idealized particle moving uniformly with the actual particle at the instant, while u, v, w are the velocity components in these coordinates. It would be possible to adhere to these same equations in the special theory, with t, x, y, z similarly defined, and a complete theory would be thereby determined. However the slightly modified form

$$\frac{\partial \rho}{\partial t} + \rho \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) = - \rho \frac{\partial g}{\partial t},$$

$$\frac{\partial u}{\partial t} = \frac{\partial g}{\partial x}, \qquad \frac{\partial v}{\partial t} = \frac{\partial g}{\partial y}, \qquad \frac{\partial w}{\partial t} = \frac{\partial g}{\partial z},$$

is superior in elegance, since all four of these equations can be embodied in a single tensor equation. This is a decisive advantage, so that we can accept these as the proper form which the theory may be expected to take.

In this way a very compact and elegant extension of the special theory of relativity arises, which takes account of gravitational as well as electromagnetic and mechanical phenomena. Unfortunately the law of gravitation so obtained will not account for the known slight anomaly in the motion of the planet Mercury. Furthermore, it possesses a disadvantage, in common with the Newtonian theory, of beginning with the straight line as the path of a particle, and then later supplying forces to account for the discrepancy caused by gravitation.

The qualitative explanation of gravitation, as far as it has been attempted here, indicates that the essential element in any new approach is likely to lie in the use of a better spatio-temporal framework. Concerning any available framework a good deal may be said in advance. In particular, if it is desired to retain anything of the older theories, it will certainly be necessary to hold to the notion that the totality of physical events corresponds to a " four-dimensional space-time continuum." This statement only means that events can be specified by means of three space numbers and one time number.

It was the mathematician Minkowski who pointed out in 1908 that, in its aspect of mathematical form, the space-time continuum of the special theory of relativity is highly analogous to ordinary space; in fact it *is* ordinary space (in the same mathematical sense as that in which the mystical equation written above holds) except that an additional dimension is present. The events are the points, the collections of events at a single particle, called *world lines*, are the straight lines, the local time elapsed between two events at a particle is the distance between them, and the space appertaining to a particle at any instant of its time is a three-dimensional "plane" perpendicular to its world line.

Now, in a sense to be explained in a moment, ordinary space is a continuum without *curvature*.

It is natural to ask in consequence: *Cannot the framework of space-time be made more flexible so as to correspond to a continuum with curvature?* The question may be restated as follows: *Cannot a space-time framework be used which bears to the framework of the special theory of relativity the same general kind of relation as a curved surface does to a plane?* It is clear that this involves a necessary requirement that each small part of the new space-time is like the space-time of the special theory in the same sense as a small piece of a curved surface is like a plane. There seems to be little doubt that Einstein's general theory of 1915 took its origin in some such line of thought as that just indicated.

A more physical statement of the central idea of the new theory may be formulated as follows. Consider an elevator cage dropping freely toward the earth. The effect of the gravitational force of the earth appears thereby completely done away with. For example, a ball thrown in the cage will describe a straight line relative to it, just as if the cage were in empty space remote from matter. This fact suggests strongly that the space and time of any one small reference particle in empty space are the same as of any other in physical properties, and that these properties have no explicit reference to gravitation. From such a point of view, gravitational effects arise because of the fact that every part of a body tends to follow a natural path, except as it comes into juxtaposition with other parts of the body.

The meaning of this interpretation may be made more apparent in another illustration. Consider a fly on a horizontal merry-go-round. The merry-go-round will have the same physiological effect on the fly as if its plane were stationary and inclined to the horizontal, the inclination being dependent upon distance from the center, and increasing with the distance. In this way a pseudo-gravitational effect is produced by the conflict between the natural motion of the merry-go-round and the natural motion of the fly.

It is evident that the relativity of physical properties to the particular particle in empty space cannot be complete. For example, in the dropping elevator cage a magnetic needle will still point toward the north. Einstein requires relativity to hold to the extent that measurements with a small clock and measuring stick can be properly made. In other words, he assumes that the familiar relations between space, time, and light hold at every small particle in empty space.

The new spatio-temporal framework is by no means completely determined by the single requirement that it is comparable to a spatial continuum. Each little part of it may be like the space-time of the special theory, in the same sense that a piece of a curved surface is like a plane, and yet a great deal of arbitrary adjustment is still possible. To effect the determination of the space-time continuum, it is necessary to discuss its intrinsic properties which, like those of any surface, hinge upon the important but technical concept of "curvature."

Consider the geometry of the plane. It may be developed independently of the fact that the plane is a part of three-dimensional space. For example, a straight line might be defined as the shortest path in it. If this kind of approach is made, a conical surface can be treated in exactly the same way as a plane, for it has precisely the same internal properties as a plane. To demonstrate the fact, it is only necessary to unroll the conical surface upon the plane. The geometry obtained indicates that the plane is similar to itself in all its parts, to the extent that, if distances are reduced by half, the plane has the same properties as before.

Now consider the geometry of the surface of the sphere in a similar spirit. The shortest line between two points is an arc of a great circle on the sphere. Furthermore, all parts of the sphere are alike just as all parts of the plane are. Angles have the same significance as in the plane.

Nevertheless, there is a characteristic difference, for the sphere is not similar to itself; for instance, if the sphere is doubled in radius, the surface of the new sphere cannot be fitted upon the old one without distortion of distances. The curvature of the sphere is measured by the extent to which the property of similarity of figures fails to hold; the plane has no curvature, and is thus flat, from this point of view.

If we consider any other arbitrary surface, like an ellipsoid, it will depart from similarity to itself at each point by an extent measured by the curvature. This curvature has nothing to do with the ellipsoid thought of as situated in three-dimensional space, but is a quantity fixed by the internal properties of the surface.

In the same way the curvature of a spatial continuum of more than two dimensions can be defined by its internal properties.

Einstein requires that space-time be as flat as possible away from matter without degenerating into the flat space-time characteristic of his special theory. This is entirely akin to the requirement that an ordinary surface not a plane should contain as many straight lines as possible. The solution of this more elementary problem is indicated in the figure shown herewith. Likewise he requires space-time to have curvature where matter is. It turns out that space-time is satisfactorily prescribed by means of these two requirements.

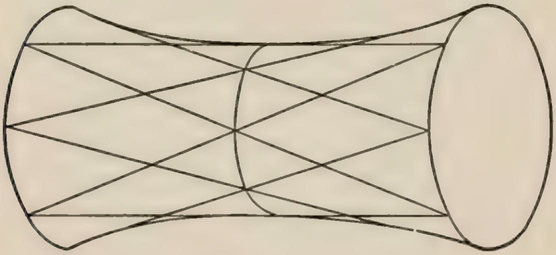

FIG. 16.

Gravitational phenomena are now explicable as the inevitable result of this curvature. More concretely, the fact that two bodies in empty space tend to approach one another is comparable to the fact that any two great circles on a spherical surface will intersect one another. The new general theory obtained explains even the anomalous gravitational behavior of the planet Mercury, and has been widely accepted on the basis of this and other experimental evidence.

It is very important to understand the essential modification which the principle of relativity undergoes in its extension to the highly fluidic space-time of general relativity. The space and time of a particle in empty space near matter are required to possess *very nearly* the same physical properties as of a particle remote from matter, so that, in particular, gravitational phenomena relatively to the particle are absent. It cannot be and is not demanded that these properties should be *exactly* the same. However, if the units of space and time selected are very small and are diminished in the same ratio, so that very rapidly running clocks and very small measuring sticks are used, a Lilliputian universe results. In this universe gravitational phenomena disappear, and the laws of space, time, matter, and electricity will be those of the special theory of relativity already considered, and the same for every reference particle.

The spatio-temporal system of any particle becomes of less and less consequence for more and more remote events. For this reason, if a well-balanced view of the physical universe is sought, it does not appear desirable to regard these individual spatio-temporal reference systems as of unusual importance. In other words, the selection of a reference system becomes to a large degree arbitrary.

It is in this fluidic property of the general theory that its difficulty resides. Near any particular event in space and time, experiments may be carried out with security at any small freely moving body; local distances, lapses of time, densities, pressures, temperatures, etc., may be treated by the usual methods, applied as if the body were at rest. The general laws of each microcosmic part will be those of the special theory, and yet there will be a slight macrocosmic deviation, caused by the presence of matter and the resultant curvature of the four-dimensional space-time continuum. Gravitational phenomena appear as only the direct manifestation of this curvature.

It is extremely natural from a philosophic point of view to require matter to condition space-time, for the physical determination of space and time is always through the intervention of matter.

A still more fluidic framework of space and time, based on the notion of "affine connection," has been introduced by Weyl, and developed much further as the "geometry of paths" by Eisenhart and Veblen. Recently it has formed the principal basis of attempts at even more complete unification of physical law by Eddington, Weyl, and Einstein.

The general relativity of Einstein is connected with his special optical relativity, the mechanical relativity of Newton, and the geometric relativity of Euclid in a very interesting way.

Suppose that we consider the general theory when matter is very sparsely distributed. Then the framework reduces to that of empty interstellar space, and the small gravitational effects noted will be those appropriate to the special theory, or will be negligible. The principles of optical relativity will hold.

Now suppose, furthermore, that matter and electricity are moving with such low relative velocities that the velocity of light may be taken as infinite by comparison. Then, according to the general theory, electromagnetic phenomena disappear, and the *régime* of the mechanical relativity of Newton sets in, and his law of gravitation will be valid.

If also the rotational velocities and internal pressures are so small that to all intents and purposes material bodies are continually in undisturbed motion, the geometric relativity of Euclid holds.

The mathematical working out of the new theory is not a simple affair, despite the simplicity and unity of the fundamental point of view which has led to it.

This is not at all different from the circumstances attending previous theories in classical physics, for example, the theory of elasticity. Thus, ordinarily the tuning fork is considered to be a rigid body. It is only in this way that it is visualized with any success. However, the very property which makes it a useful instrument is that, when struck sharply, it will vibrate and communicate its elastic vibrations to the air, and thence to the eardrums, and excite the sensation of a musical sound of definite pitch. The manner in which it vibrates is exceedingly complicated, and varies according to the way in which it is struck, although experiment shows that the pitch of the fundamental note is not thereby perceptibly altered. The determination of the exact mode of vibration of its several parts has never been made. Although theoretically possible by means of sufficiently extensive computation, provided the tuning fork is idealized as a so-called perfectly elastic body, yet it is to all intents and purposes impossible of accomplishment. It is not even possible to explain its general behavior save by elaborate theoretical considerations, nor to specify the fundamental note in advance by any formula.

And, if it were desired to deal directly with the assemblage of atoms and electric charges constituting it, the vibrational properties would appear as average effects of enormously complicated interactions of very large numbers of elements. Hence, even here, the reality is found to be so complicated as to be beyond the power of man to conceive adequately.

Simplicity and unity in the fundamental processes and yet an infinite complexity in their combinations seem more and more to be clearly manifested in nature.

Is the Einstein synthesis in any sense ultimate? That may be held to be exceedingly doubtful. The special model of interstellar space empty save for point particles, with which Einstein begins, is not a correct one, since there do not appear to exist point particles but merely smallest elements, called atomic nuclei and electrons. Without a true model as a starting point, it does not seem likely that a final conception of the physical world can be arrived at by a process of generalization. The advances in recent years in the domain of physics indicate that we are just beginning to discover the laws of the atom. In fact, we have been compelled to proceed from the evidence of our senses, and thus have observed laws which are the statistical effect of inconceivably numerous elementary processes hitherto beyond our ken.

Einstein's new theories are a great step in advance, for they bring together in a remarkable way the laws of nature so far obtained. In them, gravitation appears as the law of macrocosmic interconnection, caused through curvature of space-time by matter, while the spatio-temporal relativity of non-gravitational laws is maintained in the microcosmic domain.

CHAPTER IV

THE EXPERIMENTAL TESTS OF RELATIVITY

Thus far we have restricted attention almost entirely to the theoretical aspect of relativity in order to explain it in precise but general terms. It is now desirable for us to consider briefly how that theory has been experimentally tested.

It would be impossible to limit speculation about physical laws by the demand that a theory must apply in all respects. The vortex theory of the atom has always been considered to be of great interest, although never applied successfully. A theory of major importance will have many applications, of course, and may suggest the way to new discoveries. Thus the Newtonian formulation of gravitational law led to the discovery of the planet Neptune through the observed irregularities in the motion of Uranus.

The theories of relativity are unquestionably of very high speculative interest. The extent to which they will be found in harmony with nature is not yet fully disclosed. If a just estimate of them is desired, it is necessary to adopt the spirit of impartial inquiry in dealing with the relevant experimental facts. In this critical spirit it may be asked whether or not there is some common ground which furnishes a satisfactory basis for consideration of the questions at issue. It has been pointed out previously that there is an extensive basis of this kind.

In the first place, it is agreed that the measurement of time by more or less arbitrary means is possible, so that events can be classified by means of their time of happening, while the events happening at a particular time can be spatially classified in various ways, this last classification requiring three space measurements. An event, such as the collision of two atoms, is the elemental physical reality specified by these four measurements, and the totality of events constitutes the substratum of the physical universe now called space-time.

Furthermore, it is granted that material objects can be identified, and followed in time. Any such object occupies a definite part of space at any time. It is assumed that there is the possibility of continuing this process of identification of objects down to the point particle.

Solids, liquids, and gases may be subdivided. However, chemists, in their consideration of the various kinds of matter, have been led to the conclusion that they were made up of indivisible atoms of a definite type for each chemical element. This conception has proved of great simplifying power. More recent advances make it plain that electricity also is atomic. Magnetism is held to be merely an effect of electricity in circulatory motion. These elementary units of matter and electricity have spatial extension. Besides these, there is only space-time free from matter and electricity.

Let us turn now to the consideration of nearly empty space-time and review the generally accepted facts briefly. For the moment let us adopt the notion that a clock is available at each particle to measure the lapse of time there, and that light disturbances can be observed which originate at other similar particles. If it is assumed that such a space-time is alike in all its parts, it follows conclusively that with any such particle may be correlated an optical system of space and time, with reference to which all other particles will travel at a constant velocity in a straight line, while light travels with one and the same velocity under all circumstances.[1] The conviction that space-time, remote from material bodies, is homogeneous and of one nature throughout, has the same reasonableness as the conviction that all parts of a plane are alike. The experimental method of spatio-temporal measurement employed may be based upon a process of light-signaling between three particles relatively at rest, and the use of a clock at one of them. The space obtained by such methods will be Euclidean. Here is a considerable basis of facts about which there is no dispute.

In the spaces and times attached to small particles in empty space-time, there may happen to exist a unique one of simplest type from a physical point of view. This will be the space and time of a set of particles *at rest*. It will then be legitimate to call such a space and time *absolute*, and so we arrive at classical physics. The only alternative is that all the spaces and times are on exactly one and the same physical footing. This hypothesis yields the special theory of relativity. Which of these alternatives really holds?

---

[1] This is true because the light-second is defined as the unit of distance.

The understanding of the situation turns upon the highly difficult and interesting concept of the clock. The ordinary pendulum clock itself does not conform to what is meant; such a pendulum clock and the earth together more nearly constitute a clock in the technical sense. This particular type of gravitational clock would have to be of small dimensions for purposes of exact temporal measurement in nearly empty space-time. A pocket watch would be more nearly a clock of the desired type. Even for this elastic clock, uniform conditions of operation would be essential. There would be required constant pressure and temperature, and freedom from gravitational, electric, magnetic, or rotational stress. A body like the earth in rotation furnishes a type of rotational clock. In fact, as soon as phenomena are regarded from the proper point of view, it appears that more or less exact clocks are omnipresent in nature.

Now the statement that the universe of events is four-dimensional has no real meaning unless it is possible to identify events of a specific type T exactly. It has been assumed that particles are identifiable. The events at a particle are evidently arranged in a time order. Now if two events A and B of the particular type T happen at a particle, an interval between the two events is set off. An indefinite recurrence of such events measures the time order at this particle, and similarly at other particles, and thus the comparison of time intervals elapsing at different particles becomes possible. In other words, such a sequence as A, B functions as a clock of type T. Thus it seems clear that the mere possibility of spatio-temporal measurement involves something of the notion of the clock at any particle.

However, in both the old and new theories, the physical reality of the interval of local time is conceded. Hence it is clear that there must exist appropriate ideal types of local clocks with which this quantity can be measured. Of course the ideal clocks used must run at the same rate when together, and even when relatively at rest. There will then be physical significance in the relations between the spaces and time of various particles.

If two such clocks are separated and later brought into coincidence, it does not follow that the same length of time has necessarily elapsed on both clocks. To assume the contrary is to require the existence of absolute time.

When there is more than one kind of clock, as M and N, so that if a clock M has the same rate as N at one particle, it need not have when brought again into coincidence with N at some other particle in relative motion with respect to the first, then a difficulty arises as to whether M or N is the selected ideal type of clock. Thus, a vibrating sodium atom may be called a clock M; an oscillating spring may be taken as a clock N. If the interrelations of the various spaces and times are to have genuine physical significance, it is impossible to evade the demand that these and other kinds of clocks shall keep the same time at every particle, *unless one selects arbitrarily a particular kind of clock as ideal.*

If there is a unique physical reality, corresponding to our naïve feeling of duration at any particle in empty space, it is reasonable to suppose that all clocks will measure that duration. Nevertheless, the classical physics of Newton and Faraday has more than one type of clock. Thus an oscillating spring yields an elastic clock in the meaning of our definition. Likewise, a pair of mirrors mounted at the ends of a rod, between which light is flashing back and forth, forms a "light clock." Either of these possesses the fundamental properties required. Now will these two clocks keep the same time? The answer is in the negative, since, according to the formulation of the laws of motion, only the elastic clock will measure *real* intervals of time, while the light clock will go more slowly with respect to the elastic clock, the larger its absolute velocity. But why should either be preferred? It is difficult to find any genuine reason for the choice made.

It will be seen from this very simple example that the behavior of these two types of clocks would enable us to single out the absolute space of Newton and Faraday; it will be that space in which the light clock at rest runs most rapidly relatively to the elastic clock at rest. The corresponding time measurement can then be called absolute time. An absolute system of reference is determined.

This is not accidental. When relativity does not exist, all particles in empty space are not on a physical parity. The distinction between them will affect the types of clocks in various ways, and thus their relative rates at different particles.

The test of relativity is thus seen to be that small local clocks of any type will run at the same rate whenever and wherever brought together at a particle in empty space, provided that their rates have once been synchronized by proper choice of a unit of time, say, the second.

The test of the alternative possibility of absolute time and space is that at least two kinds of clocks can be found in nature. These will be related to one another as the elastic clock and the light clock cited above are in the space and time of Newton and Faraday. With the phrasing of the distinction in such simple terms, the immediate consideration of the facts is in order.

Here a remark must be interpolated in fairness to the classical theory. The possibility of accepting it lies in the fact that bodies found in nature have a small absolute velocity, *i.e.*, that the space attached to the sun or to any star is very nearly the absolute space. If this were not the case, the fact would be revealed in various ways, for example, by the asymmetric behavior of the spectra of the stars. Now when absolute velocities are small, it is not difficult to prove that relativity of motion will be simulated to a considerable extent. Witness the aberration of light and the Doppler effect, as explained on the old basis. These turn out to depend, as far as the principal effect to be observed is concerned, on the relative velocities only.

Hence, when we are dealing with bodies moving at low relative velocities, it is necessary to achieve a high order of accuracy in any crucial experiment. Since the ratio of stellar velocities to the velocity of light is of the order of one to one thousand, the method will have to possess an accuracy of one part in one million. There are not many experiments in which such a degree of accuracy can be obtained, but they are of great interest.

The equations of Maxwell have always been accepted as applicable to electromagnetic waves in empty space-time, where the space and time are obtained by ordinary means of measurement. The velocity of these waves can be determined theoretically and experimentally; it is independent of the frequency of the exciting electric oscillator, and is the same as the velocity of light. This last fact, and the very great parallelism between the optical and electrical behavior of the elements, extending to the equality of corresponding optical and electrical constants, indicate that the light wave is merely an electromagnetic wave of very high frequency. There must then be something in the characteristic atomic units from which these light waves are derived which is of the nature of an oscillator or atomic clock.

The wave character of light is conclusively demonstrated by the phenomenon of interference, according to which two beams of light of the same frequency may interfere to yield darkness. By this means the most exact measurements of length can be made.

In the celebrated Michelson-Morley experiment of 1887, a beam of light was divided, sent along two bars and then brought back to interference. If there had been any variation in the relative rate of the atomic clock from which the beam of the light came, and the light clock formed by the beam and the bars, with the varying velocity of the earth, there would have resulted a slight shift of the interference bands. One hundredth of the shift expected under the classical theory would have been within the reach of observation, but none was found.

The meaning of the experiment is that the light clock and the atomic clock run at the same rate. Another way of stating the meaning of the experiment is to say that the shape of any undisturbed body will always be the same in the reference space and time which belongs to it, as measured by light-signaling and the atomic clock, or by the measuring stick. From a logical point of view the experiment is corroborative but not crucial.

Another kind of clock can be defined as follows: Consider a tube filled with water and not rotating, and a beam of light which traverses the tube. It will take a certain interval of time for the beam to traverse the tube of water, as measured in the attached space and time. This will then yield a new kind of clock.

The situation may be looked at as follows: If there is an absolute space, and the tube is not at rest in it, then in the space attached to the tube, the absolute space or ether will appear to be moving through the tube in some particular direction. A beam of light which travels in the same general direction as the ether would seem likely to have a different velocity than when sent out in the opposite direction, and this difference in velocity ought to be measurable, provided that the ether is in any sense a medium as the water is. In other words, the new kind of clock would be expected to yield a measure of time, differing according to the absolute velocity of the water.

Here again no effect of the anticipated kind revealed itself under most careful duplication of the Fizeau experiment.

It is generally held that the Michelson-Morley experiment takes a central position in the experimental testing of relativity. Hence, it is of importance to see how it has been interpreted by those who hold to the old theory of space and time. The hypothesis of "ether drift" has been advanced as an explanation, according to which the earth carries the ether in its vicinity along with it. Since the original experiments were carried out in a building, such a drift was felt to be especially likely to cause the negative result.

The logical difficulties involved in the concept of ether drift cannot be too strongly emphasized. On the previous theory the ether was a particular absolute space singled out by its property that light traveled in it with the same velocity in all directions. There was now tacitly assumed to be some further underlying reference frame — the true space — with no definite physical properties.

Nevertheless, it was of great interest to repeat the experiment, preferably upon a high mountain, as D. C. Miller did at Mount Wilson in 1921. His measurements seemed to him to indicate a shift in the fringes of about one fifth that required by the classical theory.

A second effect, not expected on any theory, was also found. After apparent elimination of direct disturbances from ordinary heating and magnetic action, these two effects remained. However, the error due to radiant heat required further investigation and seemed likely to explain the two indicated small shifts observed.

Within a few months Miller has announced that extensive further experiments made by him agree with these earlier ones. It seems prudent, however, to withhold all opinion as to their actual significance until the observations are published in detail. In this connection the following statement by H. N. Russell, one of the three Research Associates of Mount Wilson Observatory, may be quoted:* "In the present case, in spite of all the careful precautions taken, the existence of such still-elusive errors appears to be more probable in the second set of experiments [*i.e.*, those by Miller at Mount Wilson]: for the very condition that they be made in as light a housing as will protect the very delicate apparatus from the disturbing influence of differences of temperature, magnetic forces, and so forth, renders them more subject to residual disturbances than the observations made in an underground chamber."

Most fortunately, Michelson and Gale have recently made a series of experiments in which a beam of light was passed around a rectangular pipe, one mile in length, from which the air had been exhausted. On the basis of either the classical or relativistic theory, the predicted shift of the interference fringes was .236

---

* Scientific American, August, 1925.

of a fringe, while the observed shift was .230, and agreed with that predicted within the limits of experimental error. This effect was caused by the earth's rotation, and would have been very much smaller if the hypothesis of ether drift were correct.

Moreover, it is equally difficult to explain the well-known facts of the aberration of light on this hypothesis of ether drift, although they are entirely in harmony with either the classical or relativistic theory.

Some physicists and astronomers, in opposing the new theories, are willing to accept the Lorentz contraction hypothesis in the interpretation of the same experiments. This is particularly difficult to understand, because Lorentz accepts the theory of relativity as the proper expression of his own hypothesis. The reason for the essential equivalence was presented earlier.

The fact is that Einstein has provided the reasonable way out of a dilemma in which physics was placed by various experiments, including those mentioned above. Without the readjustment of the concepts of space and time, to the exceedingly slight extent contemplated in the special theory, theoretical physics would be in a difficult position to-day.

The electromagnetic equations of Maxwell have a complete relativistic form to begin with, while the equations of classical mechanics can be modified slightly so as to have such a form. The behavior of an electrified particle at velocities comparable to that of light under electromagnetic forces is found experimentally to follow the law thus obtained. It can be said that the known facts of mechanics and electromagnetism are then adequately accounted for. When this is not done, the attempt to explain the facts seems to lead step by step to complete relativistic form. The obvious inference is that the easiest manner in which to grasp the physical situation is to assume the principle of relativity at the outset. If we refuse to do so, the difficulties seem endless. The immense usefulness of this spatio-temporal principle of relativity has been pointed out previously.

The argument of convenience thus becomes paramount. It is the same consideration as that which established the Copernican theory as against the Ptolemaic.

Moreover, no mean philosophical importance attaches to the fact that, if absolute space and time exist, then the interval of local time is devoid of a unique physical meaning. The theory of relativity has the unquestionable superiority that for it local time has physical reality.

The notion that the interval of local time is a basic reality may be looked upon as the guiding principle of the general theory of relativity to whose experimental verification we turn next. Its successes afford substantiation of an important kind for the special theory.

The interval of local time between events at a particle suggests the distance between points of a line to the mathematical mind. Minkowski pointed out that the space-time of the special theory was analogous to a flat surface. According to the general theory of relativity of Einstein, when matter is present in large quantities the space-time is analogous to a curved surface. It is such a curvature which is held to account for gravitation. On the basis of the hypotheses that the well-known purely mathematical measure of curvature indicated the density of matter and energy, and that empty space-time was as flat as possible, Einstein obtained a definite theory which explained the observed facts of gravitation. The new gravitational law was as uniquely determined as that of Newton. Consequently, the tests of the theory by means of the slight predicted differences may be regarded as especially conclusive. The variations from the older theory were declared by Einstein to be within the reach of observation in three cases, all of them depending upon the large mass of the sun as their cause.

Firstly, there must be a rotation of the line of symmetry of the orbits of the planets in their revolution around the sun, by an amount in excess of that demanded by the Newtonian theory. This additional effect is largest in the case of the planet Mercury. Fortunately, the angular advance of the perihelion of Mercury had been measured with accuracy. Before the theory of relativity was thought of, there had been found an advance, beyond that expected on the Newtonian basis, of about forty seconds a century. This unexplained fact had been held to be the one real difficulty in the way of a complete success of the Newtonian theory as applied to the planets.

The Einstein theory indicated an advance of the right amount for the perihelion of Mercury, and advances of the right order of magnitude for the other planets. These are of minor importance compared to that of Mercury in the opinion of astronomers generally.

When it is borne in mind that, if the advance had been of any other amount, it would not have been possible to readjust the Einstein theory to suit it, the importance of this verification is obvious.

It is possible to account for the same advance on the old basis by a suitable distribution of fine particles of matter about the sun. The most competent opinion is that there is no evidence for such matter in the amount required.

The second effect to which the theory led had to do with a hitherto unobserved phenomenon, namely, the definite bending of star light near the sun. The abandoned corpuscular theory of light of Newton led to the idea that bending might occur, due to the gravitational attraction of the corpuscles by the sun. The amount to be expected on the Newtonian basis is eighty-seven seconds. Einstein arrived at the same quantitative result by a different method in 1911. The paper of 1911 was interesting, as the forerunner of the paper on general relativity of 1915. According to the general theory, the amount of deflection to be expected is nearly double that predicted on the Newtonian basis.

In order to test the prediction, English astronomers made special expeditions to observe the total solar eclipse of May 29, 1919, and found evidence of a bending of light of about the amount indicated. The measurements require the utmost skill. The observations agreed with the Einstein theory within the limits of error (about 20 per cent.), and were held to constitute another definite confirmation of the theory. Nevertheless, it was extremely desirable that further observations should be made in order that the results be definitely established.

The first occasion for so doing was furnished by the solar eclipse of September 21, 1922. An American expedition under the direction of W. W. Campbell was successful in obtaining an excellent set of photographic plates. The plates showed a radial displacement of the star images of the magnitude demanded by the Einstein theory. The mean value from four plates gave 1.72″ as against Einstein's predicted value of 1.74″. The probable error of the plates was only .11″. This initial result was in extraordinary agreement with the theory.

The notion has been advanced that the gaseous matter which exists in the neighborhood of the sun may have caused refraction of the star light. However, our knowledge of the sun's outer neighborhood is sufficient to indicate that this is an insufficient explanation, although it may account for the slight excess found by Campbell in another subsequent set of measurements.

The third effect to be expected according to Einstein was a shift of the solar spectral lines of any element toward the red end of the spectrum, as compared with a terrestrial spectrum. This is exceedingly difficult to detect, although the general consensus of opinion among those who have investigated it, has been that the shift toward the red does exist. The latest announcement by St. John, made on the basis of new knowledge concerning the solar envelope, is favorable to the theory. It is "that three major causes are at work in producing the regular differences between solar and terrestrial wave lengths and that it is possible to disentangle their effects; namely, the slowing of the atomic clock in the sun to the amount predicted by the general theory of relativity, radial velocities of moderate cosmic magnitudes and of probable directions, and differential scattering in the longer paths traversed by the light coming from the edge of the sun."

Most fortunately again, the existence of this third Einstein effect in the spectra of certain extraordinarily dense stars has been observed at Mount Wilson. The effect is too large for any uncertainty and is of the magnitude predicted by Eddington.

All things considered, the experimental facts bear out the new theories. The results point toward the physical uniqueness and reality of the local interval of time but not of absolute time, the relativity of physical law to any undisturbed particle in empty space, with gravitation as merely the cosmic result of the direct conditioning of space-time by the presence of matter. These conclusions are of major importance for physics.

CHAPTER V

SOME RELATIVISTIC PARADOXES AND THEIR EXPLANATION

In my experience, there are a number of paradoxes which need elucidation before the feeling that there is something fundamentally wrong with the theory of relativity is conquered. It is my purpose in the present chapter to try to explain these as well as possible.

A first and insuperable paradox of the new theories is often seen in their abandonment of absolute simultaneity; it is taken to be self-evident that either two events are simultaneous or one happens before the other, and it is vigorously contended that the significance of such a statement has nothing to do with the particular method adopted for measuring time.

Very similar feelings must have been aroused when the opinion that the earth was round and not flat was first advanced, for everyday experience as well as primitive cosmogonic speculation seemed to require such flatness. In order to remove the appearance of paradox in the new hypothesis, a first step was to observe that " down " means merely " towards the earth," and need not refer to a fixed direction in space, although very nearly the same direction for long distances. Undoubtedly, a distinct mental effort was required in granting as much.

In the same way, even after the relativist points out that simultaneity obtains a genuine meaning only by reference to some system of time measurement, it is not easy to give up the belief that simultaneity is an absolute affair, and that clocks everywhere are to be adjusted so as to run in unison.

A further step was available to convince the skeptical but intelligent inquirer that the earth was spherical, namely, to inform him about the truths of geometry and the observed difference in apparent position of the sun at any two places remote from one another on the earth. Likewise, the new analysis of the concepts of space and time, and facts such as the Michelson-Morley experiment, open the way to an understanding of relativity.

The salient feature of the special theory of relativity is that all undisturbed bodies are held to be equally satisfactory for setting up systems of space-time reference. It cannot be expected that the facts observed to transpire in any particular system so obtained have an absolute physical meaning; by their very nature they must be facts affected by their relation to the arbitrarily selected reference body or "particle," as we shall call it. In consequence, there is no *a priori* basis for the identification of the facts as stated for different particles.

The apparent paradoxes of relativity are largely due to a confusion of distinct systems of space-time reference. On this account they are easily explained after suitable consideration has been given to the nature of the correlation between the various systems.

There are, of course, many results which may be stated independently of the arbitrary reference particle. For example, if light flashes back and forth along a meter stick, a perfectly definite interval of local time will elapse meanwhile on a clock attached at the end of the stick whatever its apparent velocity may be. This is the essential physical truth embodied in the so-called hypothesis of the constancy of the velocity of light, and makes it possible to measure the natural dimensions of a body by optical signaling between its points.

On the other hand, the simultaneity of two events has only a precise meaning in relation to the particular reference particle selected. Thus, if a reference particle *P* is the midpoint of a meter stick, a light flash from *P* will reach its two ends simultaneously in *P*'s system of space-time measurement, by the very definition of simultaneity. Now imagine a second stick which is moving relatively, the midpoint *Q* of which is superposed on *P* at the instant when the light flash starts. It is clear that (for *P*) the light flash will reach the approaching end of the moving stick before it reaches the receding end. In other words, the flash will not reach the ends of the moving stick simultaneously in the time system of *P*. But in the system of the midpoint *Q* of the moving stick, these two events will be simultaneous, of course.

Similarly, there is no reason to expect that the optically determined length of a meter stick in motion relative to a reference particle will be a meter which is its natural length, or that a moving clock will tick once every second in the time of a reference particle. The determination of what the facts are will depend on the selection of definite technical means for measuring the space and time of the reference particle, together with the correlation of these spaces and times in accordance with the hypothesis of relativity.

The surprising extent to which the statement of the results in the old or new theories depends on the particular technical means of measuring space and time may be brought out in a suggestive way as follows:

Suppose that a dark star enters the solar system with a velocity nearly that of light, and becomes luminous in so doing. A terrestrial observer will *see* it approach with a velocity greatly exceeding that of light. This is obvious, since if its velocity were equal to that of light, the star would strike the earth without ever becoming visible. The apparent paradox disappears, of course, when the proper allowance is made for the time which light from the luminous star takes to reach the earth.

Henceforth it is to be understood that the space and time used are not those of events as seen at the reference particle, but with this time correction allowed for.

The relativistic correlation of the spaces and times of various reference particles is very much like that to be found in some ordinary correlation problems, which can be more readily grasped, and which will be referred to first.

Suppose that the squares of a projected city are laid out, but not exactly from north to south and from east to west because of topographical difficulties. Suppose further that it is subsequently decided to run the streets differently, say inclined at an angle of 20° to the directions first chosen. It is apparent from the figure that there will be no obvious relation between distances east and north, as measured in the two arrangements.

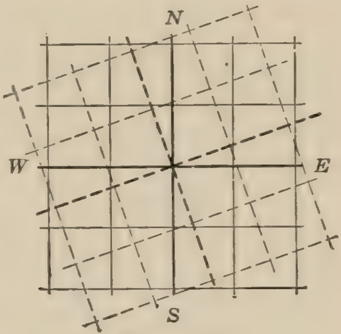

Fig. 17.

The easterly distance and the northerly distance between points of the city, as obtained in the two ways, are comparable to space distance and time difference between events for two reference particles in relative motion. Taking events which are simultaneous for one particle to be simultaneous for the other also, is akin to taking points on the same east and west street of the original arrangement as likewise so in the modified one. More generally, taking the distance or time between two events to be the same for both reference particles is like assuming that the distance by which one point is east or north of another will be the same whichever way it is measured.

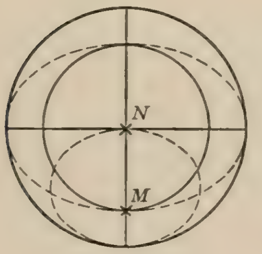

Fig. 18.

The following problem in correlation is even more analogous to that found in relativity. Suppose that the surface of the earth is mapped on a plane by some specific method. The parallels of latitude and the meridian circles will then take definite positions (the circles and the straight lines in Fig. 18). Instead of the north pole N, some other such pole can be selected, say the pole M. By the same method, a second map is obtained which will be quite different in appearance from the original map. If, however, the new parallels of latitude and circles of longitude are drawn on the first map, it will be possible to use them for drawing the second map. Here also is a problem in two-dimensional correlation.

There is no simple rule of correlation between the latitude and longitude of a point as measured on the two maps, and the fact serves once more to illustrate what may be expected when the various spatio-temporal systems of relativity are correlated.

A simple case of that theory is presented by particles moving on a single spatial line. A variety of particles are then directly approaching or receding from one another in empty space. To each of these there is attached a space and time found by light-signaling methods. Thus, if a light flash is sent out from a reference particle *A* and returns by reflection at some second particle *B* after two seconds, then the event of the reflection is specified in space and time by the two observations at *A*. The space and time referred to is that of the particle *A*; as stated previously, the accepted technique of determination of these measurements is identical with that used in the classical theory when the reference particle *A* is "absolutely at rest." Since the light flash travels the same distance either way, the distance of the event signaled from *A* is one light-second. The time of the event will be one second after the light flash was sent out from *A*.

The collection of all these events is clearly two-dimensional, with one spatial dimension and one temporal dimension.

If an " observer " at *A* were to collect observations of events, he would naturally employ a chart in which one direction indicates distance from *A*, and the perpendicular direction indicates time of happening. Such a chart serves to specify events in the same way as the plane map serves to specify the points on the earth's surface. It is to be expected that the representation of events obtained is imperfect, but so is that given by a plane map of the spherical earth. In fact, the physical reality of lapse of local time between two events at an arbitrary particle is imperfectly represented by distance in the chart, just as the physical reality of distance between two points on the surface of the earth is not represented exactly by distance on the map. For the map, the distortion is least near the poles and greatest near the equator. With the charts, a considerable distortion is found only in the case of relative velocities comparable to that of light.

Moreover, in the same way as with different choices of a pole on the earth various maps are obtained, so all the various particles yield corresponding charts. These charts will be essentially the same only for particles relatively at rest.

There is another analogy which is worthy of note. All the parts of a sphere are alike from a geometric point of view, and yet this identity of properties is masked when the sphere is mapped upon the plane. Similarly, it is a fact that space-time is the same physically for every reference particle, and yet this is not obvious from a casual inspection of a single chart. It is desirable now to outline briefly how there can be complete relativity of this kind with respect to any reference particle, without essential paradox.

Assume that a definite particle *A* is selected, and that in its space and time, determined by the particular light-signaling process specified above, every other particle travels with a constant velocity less than that of light, the latter velocity being unity on account of the fact that the light-second is taken as unit of distance in defining the space and time of the reference particle. This assumption means that the events at any observed particle *B* are represented by a straight line on *A*'s chart.

Thus far, the only clock used is the one at *A*. The clock at a particle *B* in relative motion can be arbitrarily assumed to go at any assigned rate, and a moving measuring stick to vary in length for *A* according to *B*'s velocity, and still no inconsistency can arise, for all the hypotheses are mutually independent.

Suppose in particular that any three Pythagorean numbers such as 3, 4, 5 (so that $3^2 + 4^2 = 5^2$) are selected. If we desire, the specific assumption about the behavior of clocks and measuring sticks may be made that, if the velocity in *A*'s space and time is three-fifths that of light, then the slowing down of clocks and the contraction of measuring sticks will be in the ratio of 4 to 5. We may extend this rule to any set of three Pythagorean numbers such as 5, 12, 13. In such a way there is obtained a purely hypothetical space-time, which is bound to be self-consistent by its very manner of definition.

At the reference particle B it is possible to imagine the same signaling process carried out, in order to obtain B's system of space-time reference. It will then be found as a result of easy algebra that the relation of A's space and time to B's space and time is exactly that of B's to A's. By the use of the same method it is also possible to find the interrelation of the space and time of a particle C with B's, by the intermediation of A's space and time. Exactly the same Pythagorean rule is found in this way to hold for B as well as for A. Hence the self-consistency of relativity follows. It may be observed that if clocks and measuring sticks do not behave as required by the Pythagorean rule, the principle of relativity does not hold.

The adjoining figure illustrates an optically determined position of a very long stick B of one light-second in length, and which is moving to the right at three-fifths the velocity of light relative to the reference stick A whose natural length is also one light-second. It will be observed that B's optical length is only four-fifths that of A. The clocks attached to the two ends of A point to the same zero of time, since a single instant of A's time is represented. The clock attached to the left-hand end of B is also

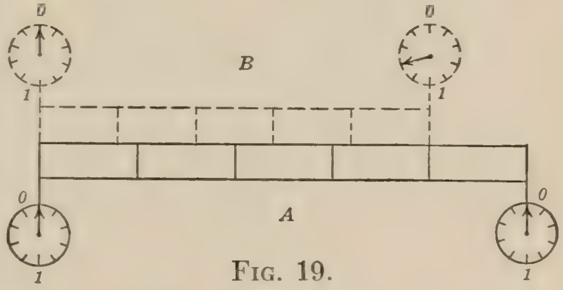

FIG. 19.

taken to point to the zero of time; it will be noted that then the clock at the right-hand end of *B* has a lag of three-fifths of a second, in agreement with the fact that events simultaneous for *A* will not be for *B*.

After one second has elapsed at *A*, the stick *B* has moved three-fifths of a light-second to the right, of course, and has still the optical length of four-fifths. It will be seen that the same characteristic lag is found in the clocks at *B*, and that these have only run four-fifths of a second for *A*.

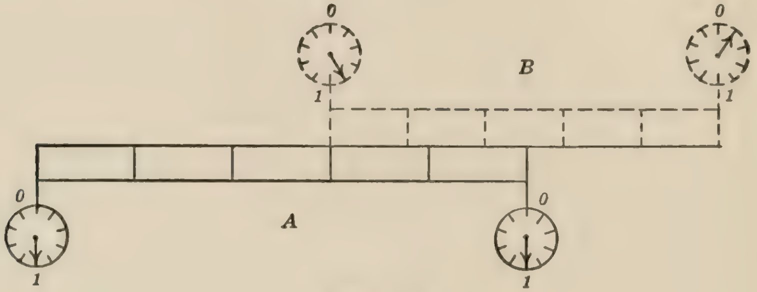

FIG. 20.

A possible difficulty may be felt in the fact that the stick *B* in motion is deemed shorter than the stick *A* at rest, while, if *B* is taken to be at rest, *A* is held shorter. Such a difficulty only arises from the attempt to overlook the difference between the reference systems attached to *A* and *B*, and treat them as identical. A glance back at the two sets of streets of the hypothetical city (Figure 17) shows that the easterly distance between a pair of north and south streets of the second arrangement, measured in the first, exceeds that between a pair of north and south streets of the first arrangement, and that, notwithstanding, the same is true when the rôle of the two arrangements is interchanged. This analogy indicates the lack of ground for the stated difficulty, or for any similar difficulty, in the behavior of bodies in relative motion.

After these preliminaries we are fully prepared to deal with some amusing and instructive relativistic paradoxes, the explanation of which illustrates various results of the theory.

The Quest of the Absolute

Despite the *dictum* of relativity, may it not be possible to define absolute space and time? The method to be proposed can be best understood after its analog in Newton's theory for determining an "absolute space" has been formulated.

According to the Newtonian laws, the center of gravity, C, of the sidereal universe moves with uniform velocity in a straight line. The "absolute space" may be taken to be that attached to a hypothetical particle at C. In other words, the absolute space is the particular one for which the center of gravity is at rest. It is for this space only that the motion of matter everywhere is most symmetrically distributed in direction, and that the total energy of motion is least.

In the special theory of relativity it is possible to proceed by a somewhat more involved but entirely similar method, to define an absolute space and time. It is simplest to do this when there is a set of mass particles in uniform rectilinear motion. Then there can be found a hypothetical reference particle C, for which the attached space and time have a peculiar property, namely, that there is no " apparent momentum " in any direction. It is again a system of reference for which the motion is symmetrically distributed in all directions. When there are only two particles of equal mass, *C* will lie halfway between them at rest in the absolute space. If the mass particles are interacting, it is still possible to obtain a hypothetical particle *C*, such that for any event *E* at *C*, some particle *D* at the event has the symmetric distribution of motion sought. Here we may define the absolute space as that of *D* at *E*, and absolute time as given by the clock moving with *C*. The main fact to observe is that, despite the increasing complexity of definition, the space and time called absolute can be chosen by a symmetric method. Even in the general theory of relativity it is possible to proceed similarly, since it is known that the space of a particle at an event can be uniquely defined.

The fundamental objection to any definition of this sort is that the most convenient statement of physical law is accomplished without any use of such an absolute system of reference. It is true that the absolute space and time specifically proposed present a further difficulty, namely, that large quantities of matter, so extremely remote from the sidereal universe as not to be observable, might greatly modify the space and time called absolute, which seems unreasonable. But mere intellectual convenience is the major consideration requiring the abandonment of an artificial absolute space and time.

The essential meaning of relativity is thus that the laws of nature are most easily stated in terms of systems of space-time reference not uniquely defined.

THE CLOCK PARADOX

The following paradox is often baffling. Two clocks K and L are at a particle P. If the second clock is moved in any manner away from P and then back again, it will be found to have gone more slowly than the first clock, although the rates will still be the same at the two comparisons. However, the situation seems entirely reciprocal. In the same process, the first clock leaves the second and is then brought back into coincidence with it. It appears as if not only L registers less time elapsed than K, but also K registers less time than L. Now such a state of affairs is obviously impossible.

The answer to the paradox is that only K remains in undisturbed motion at P. By the stated rule, it is clear that L will go more slowly than K, so that less time will elapse at L. Since the motion of L is disturbed, it cannot be taken to coincide with a reference particle during the process. If L were not disturbed it would necessarily remain in coincidence with P. Thus the reason for the apparent paradox lies in a deceptive appearance of symmetry between the two clocks.

THE AGE OF LIGHT

Astronomers tell us that light arriving from a very distant star is extraordinarily old. But, according to the theories of relativity, light is ageless. In fact, imagine a clock traveling as fast as light. The Pythagorean rule shows that the corresponding Pythagorean triplet is 1, 0, 1. Hence, contraction of measuring sticks to *nil* and a complete stopping of clocks takes place in motion with the velocity of light. Since the clock at the light wave is not going, the wave cannot age at all. More precisely, we may imagine that the clock travels at a rate slightly less than that of light; then, in the reference system attached to this clock, the aging of the light will be very small while the light leaves the distant star and travels to the earth. Of course the apparent distance from star to earth will be greatly reduced for such a reference system, as obtained by the optical method. The conclusion follows that the age of a light wave is entirely relative to the particular space and time of reference selected.

For the astronomer the reference space and time selected is that attached to the sun and the fixed stars. It is in this sense that the usual astronomical statement has meaning.

AN ADVENTURE IN TIME AND SPACE

The episode which we now propose to tell is one which seems at first to transcend the imagination, and yet which is entirely in order for the relativist.

A youth, bidding farewell to his companions, enters a properly constructed projectile and is hurled away from the earth with a velocity nearly that of light. (As a matter of fact it would require at least a month of constant high acceleration to acquire safely such an extraordinary velocity.) The projectile is directed toward an enormous star thirty light-years distant. On approaching the star, its gravitational attraction is used to reverse the direction of the projectile, and he returns safely to the earth. For him the entire journey will take a very short time and he will still be young. Yet his companions will have become old men!

This paradox is explained just as the preceding one was. Relative to the earth, his clock will tick off the seconds very slowly, and his physiological activities will go on at a proportionately slow rate. If we undertake to specify what happens relative to the daring youth, we are no longer on familiar ground, since the projectile is not in undisturbed motion during its entire flight.

It is interesting to inquire what would be "seen" to transpire through suitable telescopes. From the earth the course of events would appear to go on very slowly at the projectile on its sixty-year outward journey at apparently only half the velocity of light. Then the projectile would be seen to shoot back with an incredible velocity greatly exceeding that of light, and would reach the earth in a very short time during which events in the projectile are observed to happen very rapidly.

THE SPEED PARADOX

It has been stated that the velocity of a particle cannot exceed that of light in relativistic space-time. But why can we not proceed as follows? Consider a first particle P, moving in one direction with two-thirds the velocity of light relative to Q, and another particle R, moving in the opposite direction with the same velocity relative to Q. Then it would seem that the velocity of R relative to P is four-thirds that of light.

It is clear that this conclusion involves the assumption that the velocity of P relative to Q and the velocity of R relative to Q will combine to give the velocity of R relative to P. But what is really obtained by this addition is the velocity of R relative to P *in the* Q *system of reference,* and not in the P system, as desired. Thus, this paradox, too, is explained by the confusion of reference systems. As a matter of fact, the velocity of R relative to P will be only twelve-thirteenths of that of light. The same process may be continued indefinitely for any number of particles, each of whose velocities relative to its predecessor is two-thirds that of light; and yet the velocity of the last particle relative to the first will still be less than that of light. The mathematical formulas justify this remarkable conclusion.

IS MASS CONSTANT?

In classical physics mass is a constant. If a force, say, equal to only one pound, is constantly exerted on a pound of matter in empty space, its velocity will exceed that of light by the end of a year. But this cannot happen in the theory of relativity, since no velocity can exceed that of light. The only possible conclusion seems to be that the mass increases as the velocity does. Hence mass seems to be at once constant and variable.

To explain the paradox, let us suppose the force to have acted for only one second. The mass will possess a certain velocity (about 32 feet per second) relative to the initial reference system. Adopt now the reference system for which the mass is at rest at the instant. Since the conditions are exactly as they were at first, the velocity will be increased relatively by the same amount as before during the next second, and this will continue indefinitely.

But, as we observed in the discussion of the preceding paradox, it is not possible to make the velocity equal that of light by such a process. Consequently, the velocity will increase indefinitely, but never can become as great as that of light.

Hence in the first reference system it will appear as though the mass were responding less and less readily to the action of the force. The " apparent " mass may then be said to increase, although the " natural " mass of one pound remains unaltered.

The fact that increase of apparent mass occurs with increase of apparent velocity is often stated in the more specific but not quite exact form that increase in apparent mass is equal to the increase in kinetic energy.

Is Space Curved?

It has been pointed out that in the special theory of relativity, empty space is flat or Euclidean. It is often said that in the general theory, space is curved because of the presence of matter. What is the meaning of this statement?

A curved space is one in which experiments in spatial measurement are not in accord with ordinary geometry. For example, on the surface of the sphere, distances and angles obey different laws than in the plane. Hence the sphere is curved.

By the general theory of relativity, there is not usually to be found the stationary condition essential for comparative spatial measurement. However, the solar space attached to the sun happens to be nearly stationary, since the planets only disturb the sun slightly. It turns out that the solar space so obtained is curved, so that, for instance, the ratio of the distance around a circle with its center at the sun to its distance through is not exactly the ratio of ordinary geometry.

Of course distance may be defined anew so that the space about the sun is Euclidean, in the same way as distance on the sphere may be defined as distance on a plane map of the sphere. But such distance will not then be that determined by the use of the measuring stick in space, and lacks physical significance.

Is Space Finite?

It follows in Einstein's general theory that if the density of matter does not diminish indefinitely at great distances, then the spatial universe is finite in extent. This seems very paradoxical at first. However, if we imagine a thoroughly two-dimensional being, somehow confined to move in the surface of a very large sphere (to him like a plane), then his spatial experience will be two-dimensional just as our own experience is three-dimensional. Moreover, to him space will appear not only unbounded, but infinite, whereas in fact it is unbounded and finite. This familiar illustration is sufficient to make it conceivable how space may be unbounded, and yet finite in extent. At present the astronomical evidence does not point to the finiteness of space.

This completes the selected list of paradoxes. It has appeared that nearly all these apparent paradoxes have arisen because of a confusion of different systems of space-time reference. Yet it is possible and convenient to hold fast to a single arbitrary system of reference, provided only the arbitrariness be recognized. In the same way, position on the earth is defined once for all in terms of latitude and longitude, which involves the arbitrary initial choice of a pole of reference. Despite the fact that an arbitrary method of earth measurement is selected, no one feels obliged to consider the interrelation of two systems of reference obtained by selecting two poles.

Of course the alikeness of the sphere in all its parts and in every direction, and the fundamental importance of the concept of distance, are admitted, and the distortion of distances on the map because of the selection of a particular pole is expected.

Similarly, in dealing with a particular physical problem, it is desirable to adhere to a definite means of space-time measurement, while the alikeness of space-time throughout, as embodied in the principle of relativity, and also the fundamental nature of the interval of local time between events, are recognized. Peculiarities observed in the behavior of clocks, measuring sticks, etc., are accepted as only natural, for it is understood that these phenomena represent a kind of apparent distortion relative to the selected reference body. In this way it is easy to arrive at a satisfactory mental attitude toward the new doctrines.

CHAPTER VI

SOME GENERAL PRINCIPLES OF THE NEW PHYSICS

THERE are two kinds of ideas involved in the comprehension of a physical theory. At the one extreme lie those which are of philosophic nature, and are rich in possibilities of technical elaboration; the first formulation of the "principle of least action" in dynamics was of this type. Such germinal philosophic ideas are based on observation. They tend to be evaluated in more and more precise abstract form, until there arises finally a complete theory, which can then be set forth almost regardless of the intuitive ideas from which it sprang. Correlative with this theory there go various precise experimental processes by which the terms used acquire concrete meaning. The understanding of the completed abstraction and of the intuitive beginnings are alike of the highest importance.

In all of the preceding chapters our concern has been mainly with the intuitive ideas which lie at the basis of the new relativistic physics. With the aim of explaining more definitely how these beginnings give rise to a complete mathematical theory, there will be given in this chapter a sketch of a development of this kind, which starts from certain general principles or postulates. In the immediately following chapter these principles will be supplemented so as to lead to a perfectly definite theory of the structure of matter.

These two chapters will naturally be more intelligible to a person who has been somewhat trained in mathematical thought than to one who has not. But I believe that it will prove of advantage for all to read the explanatory non-technical material, and to glance over the equations embodying this material in the terse language of mathematical symbolism. In addition to the ordinary algebraic symbols of equality, addition, subtraction, etc., these equations involve only the notation of the differential calculus for dealing with the rates of change of quantities. Moreover, such a sum as

$$f_1 + f_2 + f_3 + f_4$$

is written in the abbreviated form f_α. In other words, when a Greek letter appears as subscript or superscript, the sum of the terms obtained by assigning all values to the Greek letter and summing is thereby understood. Subscripts and superscripts are employed in order to designate in the simplest possible way the various quantities of one and the same type.

For the lack of a better term, the principles here taken to lie at the basis of relativistic physics will be called postulates. The formal elaboration of these is by no means rigorously logical, but might be expanded so as to satisfy this demand. The names selected for the postulates are such as to suggest the corresponding basic intuitions as far as possible.

I. POSTULATE OF THE SPACE-TIME CONTINUUM. *The totality of events constitutes a four-dimensional continuum.*

The technical meaning of an event is an occurrence at a point lasting only for an instant of time. A flash of light, no matter by what means it may be specified in space and time, is such an event. The existence of such idealized events has always been presupposed by the physicist. According to the postulate, their precise specification is to be accomplished by means of four numbers. In everyday experience these might be latitude, longitude, distance above sea level, and ordinary time. But the postulate implies further that, however carefully measurements may be made, it will still be possible to use four numbers to specify any event. Thus there is involved the hypothesis of a basic physical continuum of four dimensions, called space-time, whose elements are events standing in exact correspondence with all possible sets of four numbers.

In mathematical terms this means that four variables such as

$$x_1, x_2, x_3, x_4$$

with four corresponding subscripts may be used to denote these four numbers. Thus the equations

$$x_1 = a_1, \quad x_2 = a_2, \quad x_3 = a_3, \quad x_4 = a_4$$

will specify a definite event, once a definite system of space-time measurement has been adopted. The four variables may be called the coordinates of the event.

For classical physics, the type of coordinates chosen is prescribed at the outset. One variable, say x_1, denotes the absolute time, while the other three variables x_2, x_3, x_4 give the position of the point of absolute space at which the event happens, by means of the distances of the point from three mutually perpendicular fixed planes in that space. However, it is clear that the hypothesis that such coordinates exist involves various other tacit hypotheses which we do not propose to make. Rather, as a matter of systematic technique, our coordinates will be considered as selected in an entirely arbitrary manner, and so devoid of specific physical significance. It deserves to be emphasized that this is done in order to preserve an open frame of mind, and in no way involves a relativistic hypothesis.

In mathematical language we may say that the four coordinates are to be considered as independent variables. The second postulate is concerned with the dependent variables.

II. POSTULATE OF THE FIELD. *There exist certain measurable quantities or field functions, designated as p, q, r, etc., and specified for each event as soon as the system of coordinates is chosen, which completely characterize the physical properties of the space-time continuum.*

The measurement of these quantities is of course a technical matter. Familiar measurable quantities of this type in everyday experience are density, components of force, etc.

In the older physics there were certain measurable quantities apparently not of this kind, for example, the distance between the points at which two events happen, or the difference in their times of happening. Thus the quantity measured might not be concerned with a single event and so be apparently of another type.

It is easy to show that this difference is not essential, and arises because the means of measuring space and time has been unnecessarily specialized. More specifically stated, it is possible to express the distance between two events in the classical theory entirely in terms of field functions. In fact there will be a degenerate quadratic differential form

$$dl^2 = f_{\alpha\beta}\, dx_\alpha dx_\beta$$

in which the coefficients f_{ij}, sixteen in number, are field functions, such that this form represents the squared element of distance in absolute space between an event

$$x_1,\ x_2,\ x_3,\ x_4$$

and a neighboring event

$$x_1 + dx_1, \quad x_2 + dx_2, \quad x_3 + dx_3, \quad x_4 + dx_4.$$

The distance between the two given events will then be the least number obtained by summing these small distances over any path from the one event to the other in the space-time continuum, inasmuch as the straight line furnishes the shortest distance between two points. For a different choice of coordinates, there would of course be different values of the field functions f_{ij}, but the distance number similarly obtained will be the same always.

In this way it is seen that there is no limitation in taking as fundamental only measurable quantities which are field functions. Yet the spirit of the second postulate emphasizes the primary importance of those measurable quantities which are concerned with the happenings near a point at a particular time. The postulate is in agreement with the deep-seated intuition of physicists which has led them to avoid "action at a distance" as far as possible.

III. Postulate of Causation. *From the same initial state (in any system of coordinates) the same subsequent states uniformly follow. Slight disturbances of the initial state are locally propagated.*

The first part of the postulate embodies the principle of uniformity in nature. It affirms that if the numerical values of the various field functions are known, then the subsequent values of these functions are completely determined.

With the purpose in view of giving a more precise meaning to the second part of the postulate, while avoiding purely mathematical difficulties, let us consider a purely hypothetical two-dimensional space-time. Here the universe of events possesses one space dimension and one time dimension. There are two useful ways of representing these events. One is to locate the events as happening at a point of a line, in which position is specified by means of the distance coordinate x_2, and then to imagine the time, x_1, to change. The other is to represent both coordinates in a plane diagram in which x_1 is represented by horizontal distance, and x_2 is represented by vertical distance. The second method gives us a complete representation of the space-time continuum by means of a spatial continuum of the same number of dimensions, each event being represented by a point.

Imagine a slight disturbance of the existing condition to be made at an event P. That is, the disturbance is such as to affect slightly the values of the measurable field functions only in the immediate proximity of P. The possibility of creating such a disturbance lies at the very basis of the idea of causation. The meaning of the postulate is that the effect of this slight but arbitrary disturbance will spread spatially like a wave to right and left over the hypothetical spatial line, as time elapses.

This postulate does not imply arbitrary interference with the uniformity of nature. It requires only that there do exist physical states which are the same except near a single point, and imposes a condition upon the relation of two neighboring states of this sort. Consider a stretched violin string at rest. A displacement of the string at its middle point would give a disturbance of the kind referred to, which would rapidly affect the entire string.

It is obvious that if an event Q lies on the left-hand wave out from P, while an event R lies on the left-hand wave out from Q, then R also lies on the left-hand wave from P. For if Q is the event of the left-hand wave from P reaching a certain point, the left-hand wave set up at the point and instant Q, will move to the left with the first wave.

When we pass to the representation by a plane diagram, the waves to right and left are represented by two curves through the corresponding point P. Hence the paths of all possible disturbances of this hypothetical kind will be given by two sets of curves, each of which covers the plane.

It is possible to distort this plane so that these curves become the two sets of straight lines making an angle of 45° with a straight line and lying one on each side of it. The corresponding new set of coordinates is peculiarly simple in that all the disturbance waves travel with unit velocity to right and left. The differential equations of the two types of waves are then

$$dx = dt, \quad dx = -\, dt,$$

and may be combined into the single differential equation

$$dt^2 - dx^2 = 0$$

in which the superscripts are algebraic exponents. Here t is the special time coordinate taking the place of x_1, and x is the special space coordinate taking the place of x_2.

Since no disturbance in nature is believed to travel faster than light, these causal waves are properly termed light waves. In the coordinates selected, the velocity of light to right and left is unity. If the second is the unit of time, the unit of distance is the light-second.

Transferred to general coordinates, the last equation takes the form

$$g_{\alpha\beta} dx_\alpha dx_\beta = 0,$$

where the three quantities g_{11}, $g_{12} = g_{21}$, g_{22} are determined only up to an arbitrary multiplier. Thus the ratios between pairs of them are field functions and possess physical meaning. The equation may be called the *causal equation.*

When the four-dimensional space-time is considered instead of the hypothetical two-dimensional space-time, there is obtained in like manner a condition formulated precisely as the one just set down, although the subscripts run from 1 to 4 instead of from 1 to 2. There are then ten distinct quantities g_{ij} of which the nine independent ratios have physical significance. In the space-time continuum the equation gives rise to a *causal cone* with vertex at each event.

This reasoning shows that in any physical theory for which the postulate of local causation holds, a quadratic differential form may be expected to play a central role. Various recent attempts at further developments of the theory of relativity fail to take full account of this important fact.

In the two-dimensional analog, the waves of light are also beams of light along the line. But in the case of nature, there is a distinction between a wave of light, which spreads out in all directions, and a beam, which moves along a line. However, the beam is easily definable in terms of the wave. Consider an expanding circular ripple on the surface of a pond. If a second ripple is set up at some point of the first, there will be a second expanding ripple which moves tangent to the first at a single point, and this point of tangency will move in a straight line. The first and second ripple are akin to the waves in four-dimensional space-time, while the tangent point is akin to the beam. Mathematical considerations show that there is a single beam of light in any arbitrary direction, and that all such beams move with the velocity of light.

The postulates thus far employed hold in classical physics, except that gravitation is supposed to be propagated with infinite velocity, in contradiction with the postulate of causation. Leaving aside the law of gravitation, we may say that the causal equation in the classical theory may be written in the special form

$$dt^2 - dx^2 - dy^2 - dz^2 = 0,$$

where t, x, y, and z denote the usual space and time coordinates. Inside of matter, where the velocity of light is less than 1, as ordinarily measured, this equation requires special interpretation. However, since this measured velocity is less than 1 in a gas when the disturbance velocity is certainly 1, we shall assume that the same is true for all matter.

If we accept also the account of gravitation given in that theory, the corresponding equation takes the degenerate form

$$dt^2 = 0,$$

inasmuch as the causal gravitational wave is propagated instantly.

In order to proceed further, it is convenient to examine in more detail the most simple case of space-time, namely, that devoid of matter. Such a type of space-time is nearly realized in parts of interstellar space remote from any considerable body of matter. This type of space-time will be called *isometric*, inasmuch as it is alike in all of its parts.

IV. THE ISOMETRIC POSTULATE. *All parts of isometric space-time are indistinguishable in physical prop-*

erties, and are such that the path of a particle (actual or hypothetical) is fixed by its initial position and velocity.

The notion of the particle or minute small distribution of matter has not as yet been made use of. It seems impossible to avoid the notion during the process of development of any physical theory, although, in the form here obtained, the particle will be nothing more than a one-dimensional locus in space-time with certain characteristics. In particular, if one of the physical quantities of the field has a definite value along the same line, but vanishes near by, then evidently this line may be considered as the world-line of a kind of particle. It is in some such sense as this that the idealized particle may be said to exist, namely, as a singular line of the field.

In isometric space-time the electromagnetic as well as the gravitational forces are negligible. The motion of any small stable distribution of matter and electricity, called a particle, in isometric space-time is considered to be the same as of any other particle with the same initial position and velocity. This would not be true if an exterior electromagnetic field were present. Hence the second part of the postulate cannot be regarded as valid in an empty electromagnetic field.

On the basis of the new postulate it is possible to go forward with the mathematical analysis of isometric space-time. Once again it is desirable to limit attention to the highly analogous case of a hypothetical two-dimensional space-time, in order to simplify the mathematical development as far as possible. For convenience, let us assume that the kind of coordinates used before are still employed, so that the velocity of light to right and to left is unity.

Consider any event P and a hypothetical particle at the event. That particle may stand in the same physical relation to the space-time as does every other particle at P. In other words, if any two hypothetical particles with differing velocities pass each other at an event, then it may happen that, despite the difference in velocities, each is in the same physical situation. This is the case of relativity of course, but there is nothing in the preceding set of postulates which requires such an extremely simple condition of affairs. For the present we shall suppose that we are confronted by the general case, when the two particles are not in the same situation. It will be apparent later how our conclusions apply even in the special case of relativity.

If along with the particle M at the event P, we consider a second arbitrary event Q and a particle N at Q, then clearly it must be possible to choose N so that it stands in the same relation to the space-time about Q as N does to the space-time about P. This is true because isometric space-time is essentially the same about Q and about P according to the postulate. But in setting up coordinates at a particle it is necessary to select arbitrarily a direction in the spatial line as to the right and the other as to the left. Now we propose to interpret the isometric postulate in the sense that there is no physical difference in these directions; in the case of four-dimensions an analogous interpretation is to be made as to the physical indifference of spatial direction in isometric space-time. Hence there will be two possible choices of the particle N at Q to correspond to M at P, according as M and N are assigned the same right and left directions or opposite directions. In particular if the events P and Q are made to coincide, there will correspond to M a particle N whose relation to the space-time about $P = Q$ is the same as that of N, save that right and left are interchanged. In this case we will say that M and N are homologous particles at the event.

Imagine an infinite series of hypothetical particles M at the event P, whose velocities range from the extreme velocity of light to the left to that of light to the right, increasing from one velocity to the other. The homologous particles N must be arranged in a similar series in which the velocity decreases from the velocity of light to the right to the velocity of light to the left. Of course these two homologous series are entirely hypothetical in character. There will be one intermediate velocity which is the same for M as for N, so that the corresponding particle is homologous to itself. Such a particle has the remarkable property that the space-time at P has the same relation to it, whatever convention as to right and left is adopted. Evidently a particle in this self-homologous condition at an event P will continue so indefinitely. We shall call it a self-homologous particle in consequence. Through each point of the plane representing the space-time continuum of two dimensions, there will then be a straight line or a curve representing the world line of the self-homologous particle at the corresponding event.

Our next step will be to use such a particle as reference particle for determining a system of coordinates which belongs to it. Let us begin by observing how the notion of local time at such self-homologous particles is involved in the principles already accepted. If P' is an event near to P at the self-homologous particle M, then there is evidently one and only one event Q' near to Q at the world line of another such particle N, which stands in the same relation to Q as P' does to P. It is to be noted that, because of the alikeness of space-time in all of its parts, the self-homologous particles M and N at P and Q, respectively, are in precisely the same physical condition relative to the space-time. If we select a second event P'' at M, so that P'' is related to P' as P' is to P, then we may say that the interval of local time at M between P and P' is the same as that between P' and P'', or between Q and Q' at N. Thus a process of comparing time intervals at all self-homologous particles arises which gives a perfectly definite system of time measurement, once the unit of time (say the second) has been selected. Any instrument by the aid of which this physically significant quantity can be measured will be called a clock, and the quantity measured will be termed local time at the particle.

By the aid of light-signaling and such a clock at the arbitrary self-homologous particle M, it is possible to set up the desired system of space-time measurement, relative to M. Let t_1 denote the local time on the clock at which a light wave is sent out from M to the right, and t_2 the time when the reflected wave to the left returns after meeting another hypothetical particle at an event Q. Then, according to the classical technique for locating Q when M is absolutely at rest, the distance of Q in light-seconds to the right is given by

$$x = \frac{t_2 - t_1}{2},$$

and the time of Q is

$$t = \frac{t_2 + t_1}{2}$$

seconds after the epoch. By definition we adopt these as the space and time coordinates of Q for the self-homologous particle M. A little consideration shows that, owing to the manner of definition, light travels with unit velocity to right and left as measured in these coordinates, so that the new coordinates are of the special type which we have found useful.

In these coordinates relative to M, let us consider the nature of the motion of any other self-homologous particle such as N, which at any particular instant (in M's coordinates) lies at a definite distance from M. If N is either approaching or receding from M at this instant, then obviously the space-time is not the same in its nature when the sense in which time is thought of as elapsing is reversed. The meaning of complete reversion is manifest, for light waves, particles, and clocks may be taken to operate in the reverse order. The requirement of such reversibility in time has always been imposed in physics. Just as the requirement was imposed that the direction along the space axis is indifferent, so now we require that the direction of increasing time along the time axis is indifferent.

These conditions are to be regarded as part of the isometric postulate.

Hence we can infer that N is at rest with respect to M at this instant, and so always. Therefore, the self-homologous particles are merely those relatively at rest with respect to the particular self-homologous particle M, in M's coordinates. Furthermore, if light flashes back and forth between two self-homologous particles relatively at rest, the interval of time elapsed at either between successive flashes must be the same on the clocks at both particles. From these facts it follows that the coordinates of N differ from those of M merely by constants.

Thus, in the general case, an absolute space and an absolute time underlie the isometric space-time with which we started; these are the space and time common to all of the self-homologous reference particles. In the corresponding coordinates t, x, the differential quadratic expression

$$ds^2 = dt^2 - dx^2$$

has then a meaning independent of the particle with which we begin, and so measures a physically significant quantity between any two near-by events t, x, and $t + dt$, $x + dx$. We shall term this quantity ds the element of local time elapsing at a particle which moves so as to contain both events. This definition is consistent with our definition of local time in the case of self-homologous particles, for then $dx = 0$ and $ds = dt$. There must of course be a type of physical clock which measures this particular type of local time.

In the coordinates of the set of self-homologous particles, the acceleration of any particle not of this type is clearly a function of its velocity, *i.e.*, a differential equation holds,

$$\frac{d^2x}{dt^2} = f\left(\frac{dx}{dt}\right).$$

This equation must be unaltered if x is changed to $-x$ and also if t is changed to $-t$, in accordance with the reversibility of space and time. But it cannot be so unaltered unless f vanishes. Hence the acceleration of such a particle vanishes always, and its motion will be with constant velocity to left or to right.

Along a light wave we have $dx = \pm dt$, so that ds vanishes. Hence, if a particle moves with the speed of light, local time (as defined) does not elapse at the particle.

Thus far we have assumed that not all of the hypothetical particles at an event are related to their environment in exactly the same way, and we have been able to arrive at the conclusion that an underlying absolute space and time exist. It is conceivable, however, that all particles stand on exactly the same physical basis. In fact such is the situation called for in relativity.

This special case can be treated by methods which are not very different from those used above. In the first place every particle will be self-homologous. Hence, local time can be defined in the same way for all of them. Furthermore, there can be set up by use of clocks and light signals a set of coordinates relative to any assigned particle. As before, these coordinate systems will have the property that light travels with unit velocity in either direction in any such system.

Choose now one such system. The acceleration of any other particle with prescribed velocity at any definite event is then determinate, in accordance with the latter part of the isometric postulate. This fact is expressed in a differential equation of the second order,

$$\frac{d^2x}{dt^2} = f\left(x, \frac{dx}{dt}\right),$$

where the variable *t* does not appear on the right-hand side inasmuch as the law must be independent of the epoch from which time is measured at the reference particle. Unless there exists an absolute unit of time and of distance in the isometric space-time, this differential equation must be unaltered in form when these units are changed in the same ratio, *i.e.*, when the second and the light-second are changed together. Mathematically, this means that the right-hand member, like the left-hand member, is divided by k when t and x are changed to kt and kx respectively. Hence the equation is of the more restricted form

$$\frac{d^2x}{dt^2} = \frac{1}{x} f\left(\frac{dx}{dt}\right).$$

However, this yields a singularity when the particle coincides with the reference particle at an event ($x = 0$), unless f vanishes identically for all values of its argument. To avoid this difficulty we must require f to vanish. Hence the apparent acceleration of any particle vanishes always, and the motion of any particle in the space and time of the selected reference particle is with constant velocity. Consequently, in the space-time diagram the world lines of all particles are represented by straight lines. This was true in the general case also.

Thus, we can assert that there exist hypothetical particles at any distance from a given reference particle which remain at rest with respect to it. The coordinates of all such particles are essentially the same, just as they were for the self-homologous particles in the general case. The expression for ds will have a significance for such a set of reference particles. There is no obvious necessity that the ds so defined is the same as the measured local time.

The relation between the coordinates of particles M and M', however, is easily shown to be linear, since straight lines in one representation correspond to straight lines in the other. Also, the straight lines with slope 1 correspond, since these represent light waves. Hence it follows that

$$t' = \alpha t + \beta x,\ x' = \beta t + \alpha x.$$

By solving for t and x, this gives

$$t = \frac{\alpha t' - \beta x'}{\alpha^2 - \beta^2},\ x = \frac{-\beta t' + \alpha x'}{\alpha^2 - \beta^2}.$$

If we compute the velocity of the second particle ($x' = 0$) relative to the first, we find the numerical value $-\beta/\alpha$, while for the first particle relative to the second it is β/α. If the relationship is to be reciprocal, as stated, we have necessarily $\alpha^2 - \beta^2 = 1$, and thence

$$dt'^2 - dx'^2 = dt^2 - dx^2.$$

Hence ds has the same physical significance always and must be the true element of local time, since it is for the reference particle chosen.

In this way, the preceding postulate yields two alternatives. In the first there is an absolute space and time, and the formula for ds yields one kind of local time. But t furnishes another measurable type of time. In the second relativistic case all particles are on a complete parity relative to the space-time; the same ds measures local time in the only proper sense for all particles. The first possibility is eliminated by the next postulate.

V. POSTULATE OF LOCAL TIME. *All local clocks run at the same rate when compared at any particle.*

If this postulate does not hold, two clocks which are in agreement at a particle M may go at a different rate at a particle N. The question then arises as to which clock really measures the true interval of local time, and of course no satisfactory answer to this question can be given. But in the classical case there were defined two types of time, measured by ds and dt. Hence this possibility is excluded by the new postulate, and we are led to the special theory of relativity. Thus, the preceding postulate is the one which requires relativity. We might have expected this effect of the postulate, for, if there were an underlying absolute space and time, it would be natural that motion in this space would affect the mechanisms of different local clocks differently.

Hence, for three-dimensional space the lapse of local time is given by the equation

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2,$$

where the coordinates selected are those obtained by light-signaling methods. In such coordinate systems all particles move with constant velocities in straight lines, and light beams do the same but with unit velocity.

Let us consider next a type of space-time which follows the isometric case in simplicity, namely that furnished by the gravitational field in empty space. (Here only fields are considered in which there are no electromagnetic disturbances.) In such a field the isometric postulate cannot be satisfied, for it is impossible to obtain coordinates of the special type t, x, y, z just referred to. In fact, when light-signaling methods are adopted for setting up coordinate systems, bodies like the earth are found to follow curved paths and not straight lines. These divergences have been ascribed to gravitational forces; in other words a correction has been superposed on the fictitious straight lines required by the postulational development in isometric space. It is preferable to devise postulates which take account of the physical facts at the outset. A fundamental approach of this sort is furnished by the principle of equivalence of Einstein, which may be stated in the following form.

VI. Postulate of Equivalence. *In properly taken coordinates the gravitational field about an event is equivalent to isometric space-time.*

This postulate will be developed in what seems to be the natural manner. The postulate itself is strongly suggested by the behavior of a body moving freely in a gravitational field. The case cited in Chapter III of the freely falling elevator, relative to which gravitational forces disappear, illustrates the equivalence specified.

In the first place we can expect that in properly taken coordinates, t, x, y, z at a single event we shall have

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2,$$

where ds is the element of local time at any particle passing through the event under consideration. Translated in terms of general coordinates, this means that there exists a quadratic differential form

$$ds^2 = g_{\alpha\beta} dx_\alpha dx_\beta,$$

having physical significance. This conclusion is of vital significance for the mathematical development, for it implies that the machinery of the tensor calculus can be employed.

Secondly, it is natural to assume that, for ds so defined, the relation $ds = 0$ holds along the light wave, as it does in the isometric case.

Finally, this postulate is understood to imply that, if these special coordinates t, x, y, z are so chosen that all the coefficients g_{ij} are stationary at the event under consideration and so as nearly constant as possible, the particles are without acceleration at the event. This implies that the world lines of particles are geodesic (*i.e.*, shortest or longest) lines in the space-time.

It remains to formulate the condition that the differential form is one that actually belongs to an empty gravitational field. We shall do this by means of the following postulate.

VII. POSTULATE OF THE GRAVITATIONAL FIELD. *The field conditions imposed upon the functions* g_{ij} *in order that they may belong to an empty gravitational field do not depend upon their derivatives of higher than the second order.*

It is desirable to develop the meaning of this postulate. There can be no equations involving the g_{ij}'s and their derivatives of only the first order, which describe an intrinsic property of the field, for, by proper choice of coordinates, the g_{ij}'s and their first derivatives can be made to take arbitrary values. But there are such equations of the second order, of which the simplest are linear in these second derivatives.

There are known to be essentially only three types of such linear conditions.[1] If R_{ijkl} denotes the so-called Riemannian curvature tensor,

$$R_{ijkl} = \frac{1}{2}\frac{\partial^2 g_{il}}{\partial x_j \partial x_k} + \ . \ . \ . \ ,$$

---

[1] See Chapter XIII of my book, *Relativity and Modern Physics.* Cambridge, 1923.

and if R_{ij} and R denote its first and second contracted tensors respectively, these conditions are

$$R = 0, \quad R_{ij} = 0, \quad R_{ijkl} = 0.$$

The first in no way determines the field functions g_{ij}; the last requires the space-time to be isometric and so is too restrictive. The intermediate case effects precisely the desired kind of determination of the functions g_{ij}.

This intermediate form of condition can be proved to be essentially equivalent to the field conditions of gravitation selected by Einstein:

$$R_{ij} - \tfrac{1}{2} R g_{ij} = 0.$$

With a high degree of approximation these conditions yield the familiar gravitational law of Newton.

The gravitational theory of relativity thus obtained is evidently a theory of empty space rather than of space where matter is. According to it, gravitational effects are propagated with the speed of light.

Let us see how we are enabled to proceed to formulate certain delicate experimental tests without making any further assumptions. These tests are to be carried out in the strong gravitational field of the sun.

The solar gravitational field seems to possess radial symmetry about the sun. Let the equally reasonable hypothesis be granted also that solar space-time is substantially isometric at a very great distance away. A definite formula for ds^2,

$$ds^2 = \left(1 - \frac{2\,m}{r}\right)dt^2 - \frac{dr^2}{1 - \dfrac{2\,m}{r}} - r^2(d\theta^2 + \sin^2\theta\, d\phi^2),$$

(r, radial distance; ϕ, solar longitude; θ, solar colatitude) then follows. Only a single arbitrary constant m is involved, and it represents the mass of the sun. By means of the formula certain delicate quantitative tests of the new theory can be made. The advance of perihelion of the planet Mercury, the bending of star light near the sun, and the shift toward the red in the solar spectrum confirm the new theory. (See Chapter IV.)

Thus the relativistic theory of gravitation follows from a few plausible physical principles which are in no way dependent upon any particular hypotheses concerning the structure of matter. Its agreement with the facts has been so complete as to lead to general acceptance of the theory.

The last postulate which we propose to include with the general principles may be formulated as follows:

VIII. POSTULATE OF STABILITY. *There exist physical states which are of static type, and which are stable, so that small disturbances remain small for all time.*

The earlier postulates require that the laws of nature can be expressed by means of certain kinds of differential equations; the new postulate gives further information concerning the structure of these equations. To begin with, it demands the existence of a particular choice of coordinates x_i, and of corresponding field functions, $p, q, r \ldots$, such that none of the physical quantities change with the time x_1. These will correspond to a possible physical state of static type.

The idea of the possible states of a physical system is an interesting one, deserving of special remark. In any set of partial differential equations, some of the dependent variables can be taken arbitrarily. For example, the initial form and velocity of a stretched elastic string can be prescribed at pleasure; this fact is obvious as soon as the differential equation of the stretched elastic string is written down. All such infinitely many possible states of a physical system are in harmony with the differential equations which control it. The actual order of nature appears as the development of one of the possible states.

The essential arbitrary elements in a system of partial differential equations in four independent variables, such as we have to deal with, will be arbitrary functions of three variables. For instance, such arbitrary elements are provided by the initial densities and velocities of matter and electricity. Thus, it is inevitable that the initial distribution of matter and electricity making up the atom is arbitrary, as far as the differential equations are concerned. Of course, to the greatest possible extent the original state must be fixed by means of auxiliary "boundary conditions." But there will still necessarily remain arbitrary elements, such as those determining the distribution of the atoms in space.

Since the differential equations cannot involve the time x_1 explicitly, because of the principle of uniformity, it follows that when the time variable x_1 is suppressed, there will remain a set of differential equations in the dependent variables p, q, r, . . . , but in only three independent variables x_2, x_3, x_4. There will be as many equations as there are unknown quantities. Therefore, static states are to be expected, even on the basis of the former postulates alone. By continuing further in the same direction, it is obvious that there are likely to exist static states dependent upon even fewer space variables. It is these particularly uniform states which the physicist treats in his experimental work. Isometric space-time gives the completely uniform condition, in which it seems obvious that all of the physical quantities p, q, . . . can be regarded as mere constants.

Now certain nearly static states found in nature seem to be maintained permanently, despite the fact that they are subjected to minor disturbances. Isometric space-time is continually disturbed by the passage of light waves through it, but the ensuing modification is altogether transient in effect. The atoms and electrons are endowed with a like permanency. The new postulate requires not only the possibility of static states, but also their stability in this sense. In order to arrive at a well-founded opinion as to the real condition involved, it is essential first to consider the condition for stability as it has been developed in classical physics.

Take first the dynamical system with a " finite number of degrees of freedom." Here there are usually certain coordinates,

$$p_1, p_2, \ . \ . \ . \ p_n,$$

defining position, and certain coordinates,

$$q_1, q_2, \ . \ . \ . \ q_n,$$

which determine the velocities. By a proper choice of such variables and a proper combination of the mathematical equations expressing the physical facts, the differential equation of the system can be given the " canonical Hamiltonian form "

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \qquad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}$$

A value independent of any ordinary physical principle has been attributed to these equations, because they are derivable from a single " minimum principle "

$$\int (\Sigma p_i q_i' - H)dt = \text{minimum.}$$

The function H under the integral sign stands for the total energy, so that at first sight it seems to be a genuine physical principle that is involved. But it is necessary that the energy H be expressed in a certain arbitrary manner if the proper equations are to result. If we write $\partial H/\partial p_i$ for q_i' under the integral sign (a correct process), the same variation yields the wrong equations. Thus there is a concealed artificiality in the " minimum principle."

In my opinion, the importance of the Hamiltonian form of equations for dynamical systems with a finite or infinite number of degrees of freedom lies solely in the fact that the very delicate conditions for stability, first completely enunciated by Poincaré,[1] hold for these equations provided that obviously necessary inequalities are satisfied by the periodic motion under consideration. A variety of minimum principles that may be formulated have the common characteristic that they lead to systems of differential equations possessing the same type of stability, whether or not the equations have a physical origin.

---

[1] *Les Méthodes Nouvelles de la Mécanique Céleste*, vol. 1.

However, in the highly fluidic space-time which we are considering, it is merely stability about the static state which it is necessary to demand. In order that this condition shall obtain in the simpler case of a dynamical system with a finite number of degrees of freedom, considered above, it is by no means requisite that the equations be derived from some minimum principle. Thus a system of differential equations of the very general type

$$\frac{dx_i}{dt} = X_i(x_1, x_2, \cdots, x_n),$$

will possess such stability about the equilibrium state

$$x_1 = x_2 = \cdots = x_n = 0,$$

if only there is a function F such that

$$X_1 \frac{\partial F}{\partial x_1} + X_2 \frac{\partial F}{\partial x_2} + \cdots + X_n \frac{\partial F}{\partial x_n} \equiv 0,$$

where in the series expansion for F,

$$F = a_{11}x_1^2 + 2\,a_{12}x_1x_2 + \cdots + a_{nn}x_n^2 + a_{111}x_1^3 + \cdots,$$

the quadratic terms constitute a "positive, definite quadratic form." In this case F = constant is an integral. Evidently such equations do not flow in general from a minimum principle.

With these facts in mind let us ask again what the requirement of stability means in relativistic space-time.

It is a consequence of the previous assumptions that the differential equations are of tensor form and that the reversibility of the order of time is possible. If then there are slight disturbances of a static state which result in subsequent large variations, there must be other slight disturbances in which the variations tend to disappear.

When a nearly static state is at hand, however, the amount of disturbance may be measurable as " energy " which will be less, the more the static state is approximated to; and this energy may be dissipated as electromagnetic radiation outward in empty space.

Hence if the laws laid down are such that, at low velocities of matter and electricity, these indicate the dissipation of the energy of a slight disturbance as electromagnetic radiation, the desired stability will be sufficiently ensured.

Under special conditions the rate of dissipation may be exceedingly slow. Such a condition is that in which the laws of classical dynamics may be expected to hold approximately.

CHAPTER VII

THE STRUCTURE OF MATTER IN THE NEW PHYSICS

VARIOUS points of view may be taken in the endeavor to construct a satisfactory account of the detailed structure of matter. All the experimental evidence shows only about one hundred kinds of elements distributed throughout the physical universe, these elements themselves being in intimate relationships, as is manifested in the so-called periodic table. In particular, the hydrogen nucleus, carrying a positive charge, and the electron, with an equal negative charge, have been conjectured to be basic constituents. It is these alone which, arranged in the various possible stable configurations, are held to form the atoms of the known chemical elements. Ordinary ponderable matter would then consist of extraordinarily large numbers of such atomic or molecular parts. Apparently the known facts can be satisfactorily accounted for only in this way.

The account to be given here has developed in an attempt to elaborate a theory of the structure of matter which possesses great mathematical simplicity. It may be worthy of mention that a single final tensor equation contains the theory proposed. In order really to test it, considerably more elaboration needs to be made than is attempted here.

In conformity with the general scheme of approach begun in the preceding chapter, in which we passed from entirely empty space to the gravitational field in empty space near matter, let us take up another idealized state which may be compatible with nature.

In the first place let us suppose that there is a swarm of fine particles, all of which are identical physically, and so remote from one another as to collide very seldom, so that there is no pressure effect. Such a situation might be realized in a nebula. It will be assumed that these particles are so minute as not to disturb one another gravitationally, nor to affect the isometric nature of the space-time to any appreciable extent. Hence these particles will move with uniform velocity in a straight line in the coordinates obtained by light-signaling methods.

With the coordinates belonging to any one of these particles, the number of other particles in a unit cube of which it is at the center, is proportional to what may be termed the density of the distribution. Let us denote this quantity in proper units by ρ. It is approximately a field function of the type admitted by the second postulate of last chapter, and is invariant under a change of coordinates, for ρ has definite physical meaning.

On the other hand the velocity components of any particle are given by

$$\frac{dx_2}{dx_1}, \qquad \frac{dx_3}{dx_1}, \qquad \frac{dx_4}{dx_1},$$

or equally by the first order velocity tensor

$$u^i = \frac{dx_i}{ds}.$$

These functions u^i are four additional field functions.

In the idealized case which is approached as the particles are distributed more and more densely, while the particles themselves become of extremely minute dimensions, there arises a continuous distribution of "free matter." It is obvious that the laws of the distribution will be that every particle proceeds with uniform velocity in a straight line, while the mass (*i.e.*, product of volume and density) of any part remains invariable. To formulate these requirements in tensor form we need merely define the tensor of the second order

$$T^{ij} = \rho u^i u^j,$$

and write

$$\frac{\partial T^{i\alpha}}{\partial \xi_\alpha} = 0.^{1}$$

This gives us four "conservation" equations which express the desired facts, independently of the coordinates chosen. If, in particular, the coordinates are those of light-signaling type belonging to a particle, these conditions imply, respectively, for $i = 1, 2, 3, 4$, the conservation of mass, and uniform velocity components in the direction of the three axes.

It is easy to develop a satisfactory theory of the behavior of free matter in the induced gravitational field by means of the equivalence postulate. By that postulate, the equations stated above will maintain

---

[1] For the notation of covariant differentiation here employed, see my book *Relativity and Modern Physics*, p. 134.

their form in this case also. On the other hand, the field equations of last chapter have now to be altered since space-time is no longer empty. If we write arbitrarily

$$R_{ij} - \tfrac{1}{2} R g_{ij} = \chi T_{ij}, \quad (\chi = -8\pi),$$

we see that $\rho = 0$ and $T_{ij} = 0$, in empty space, so that the correct field equations hold. Moreover, for nearly stationary free matter the geodesic motion turns out to be in essential agreement with the Newtonian law of attraction. Of course it is known that this law is very nearly valid. From these new equations of the gravitational field follow the four equations first written. Hence the quantities ρ and u^i are determined to the precise extent required. In fact, there are essentially ten field equations and fifteen variables; one further relation is known to exist identically, namely

$$g_{\alpha\beta} u^\alpha u^\beta = 1,$$

and this leaves 4 arbitrary functions in the general solution, which is appropriate to the four arbitrary coordinates at our disposal.

It is possible to simplify further the complete equations for free matter in the gravitational field. If we define a tensor

$$v^i = \sqrt{\rho} u^i,$$

it is obvious that v^i determines ρ and u^i as follows:

$$\rho = g_{\alpha\beta} v^\alpha v^\beta, \qquad u^i = \frac{v^i}{\sqrt{\rho}},$$

while the relation $g_{\alpha\beta}u^{\alpha}u^{\beta} = 1$ appears as an identity. Hence we may replace the above equations by the single tensor condition

$$R_{ij} - \tfrac{1}{2} R g_{ij} = \chi v_i v_j,$$

and define ρ, u^i subsequently as above.

Now from these equations follow the 6 determinant conditions

$$\begin{vmatrix} R_{ii} - \tfrac{1}{2} R g_{ii}, & R_{ij} - \tfrac{1}{2} R g_{ij} \\ R_{ji} - \tfrac{1}{2} R g_{ji}, & R_{jj} - \tfrac{1}{2} R g_{jj} \end{vmatrix} = 0,$$

and conversely. Thus we have eliminated v^i. The complete theory now depends on these equations in the g_{ij}'s only. This is interesting, as showing that the g_{ij}'s may be regarded as the only significant field functions in a space-time containing free matter.

The theory of matter later given will contain four functions ϕ_i in addition to the ten functions g_{ij}.

That physical condition which we have called free matter is at most a particular or limiting form of what may exist in nature. Nevertheless it plays a part of some importance in the transition from the most simple condition of all, which we termed isometric space, to the general condition. The direction of generalization is determined of course by what are conceived to be the physical facts, but these are taken in a highly qualitative way. Free matter represents only an idealized physical condition, because other quantities besides those giving density and velocity are important; in particular the pressure must be taken into account.

We are going to deal next with the electrically neutral condition, which we shall call the fluid. The neutral condition throughout solids, liquids, and gases apparently sets in wherever there are no large rapid interchanges of energy. However, all matter in its atomic parts seems to be electrified, so that the neutral condition too is at best a limiting condition.

It might be supposed that free matter is the carrier of positive and negative electricity. There would be no difficulty in writing down a set of tensor equations embodying this hypothesis. However, in such a condition of matter and electricity, there would exist no stable nucleus or electron. This fact is obvious since electricity of positive or negative sign repels other electricity of the same sign, and the forces of repulsion called into play in the process are great out of all proportion to the gravitational attractive forces between the material carriers, so that rapid expansion of the atoms would take place. It would seem inevitable that there should exist a negative pressure or tension inside of the atoms which provides a counterbalancing force. It is for this reason that the simplest possible fluid which possesses such a tension will be employed as the carrier of electricity. Physical theorists like Poincaré who have aimed at a mathematically complete statement of physical laws, holding within matter as well as outside of it, have found it necessary to postulate forces of this character.

It may be asked what the meaning is of the forces of which mention is made so freely. In ordinary experience there is an intuitive interpretation of course, such as make up all our physical ideas. But, in the final analysis, the significant fact is that the acceleration of matter is found to be a sum of certain constituent parts. These parts, when multiplied by the density of matter, are the forces referred to.

With this imperfect justification of the next step in the approach here adopted, we proceed to formulate the first of the three special postulates upon which our theory of the structure of matter will be based.

IX. POSTULATE OF THE PERFECT FLUID. *The fluid state of matter depends upon a field function ρ, the density and also upon a functionally related pressure p. The fluid moves as if "free" provided that the pressure is constant.*

The special relation between pressure and density may be written

$$p = f(\rho),$$

where the function $f(\rho)$ is not specified for the present. If it remained undetermined throughout, this equation might properly be termed a "constitutive equation," and would epitomize a gap in the theoretical development.

According to the second part of the postulate, when pressure and density also are constant, equilibrium is possible just as it is for free matter in isometric space-time. This means then that constant ρ implies the conservation equations earlier written for free matter. This fact is most simply expressed mathematically by writing

$$\frac{\partial T_i^a}{\partial \xi_a} = \frac{\partial p}{\partial \xi_i},$$

where the unit of pressure is to be properly taken. If, however, the definition of the energy tensor of the fluid, T_{ij}, be extended so that

$$T^{ij} = \rho u^i u^j - p g^{ij},$$

then the field equations and the conservation equations will have the same form as for free matter. Thus the fluid defined appears to be a proper generalization of free matter. The particular properties of the fluid will of course depend upon the constitutive equation. When $p = 0$, it is free matter; when $p = \rho$, the fluid is incompressible; when $p = \frac{1}{3}\rho$, it is that particular type of fluid in which the compressional wave travels with the velocity of light in every direction.

The difference between those parts of space where the "perfect fluid" is, and those parts where it is not, needs to be mentioned. The functions p and ρ are to be regarded as defined in empty space, and as vanishing there. On the other hand, the velocity must be regarded as indeterminate in empty space, so that the velocity components are not field functions, properly speaking. Hence the equations become meaningless where there is no fluid, and we require two types of equations, the one specified above, and the other obtained by replacing the energy tensor T_{ij} by 0. In addition there must be certain boundary conditions which specify the behavior of the functions involved at the boundary of a portion of perfect fluid. Such a condition, which has always been accepted and which can be mathematically justified, is that the pressure and density of the fluid both vanish at the boundary, so that these functions are continuous everywhere.

In the later developments we shall regard the perfect fluid as carrying electricity, and so subject to electrical forces. The conservation equations in isometric space-time then take the form

$$\frac{\partial T^{i\alpha}}{\partial \xi_\alpha} = f^i.$$

In such a case we shall define f^i as the force tensor, and shall measure the work done by these forces as

$$\int f^1 dv \quad (dv, \text{ the element of four-dimensional volume}).$$

It is not difficult to show that the principle of conservation of energy will obtain if the density of static energy is $\rho - p$. This principle asserts that the work done by the forces manifests itself as static and kinetic energy. The principle will be useful to us later, since, by properly defining the density of energy of electricity, the principle of conservation of energy will hold for matter and electricity combined. It will then follow from the known laws of radiation of electromagnetic energy that only those static forms of distribution of matter and electricity can be stable which possess the minimum possible energy. In other words it will be by means of the concept of energy that we shall try to show that the specific theory proposed may have the property of stability demanded by the last of the general postulates.

The second of the three special postulates is concerned with the properties of electricity. Anyone who has done experimental work in electricity thinks of electricity as a highly mobile substance of two opposing types (positive and negative), always attached to matter in amounts which tend to exactly neutralize one another. Furthermore, charges of like type repel one another, while those of unlike type attract, according to the inverse square law. Electricity in circulatory motion is found to behave precisely as does the magnet, and brings into play magnetic forces in addition to the electric forces mentioned. Thus magnetic forces are thought of as ultimately due to closed currents of electricity.

It is obvious that the measurement of these electromagnetic forces is always accomplished through the use of matter. But it is customary and convenient to leave the technical specification of these forces aside as not a part of electrical theory in itself. It is this abstraction which makes the notion of force important.

In the past there has been only one set of equations for the electromagnetic forces which has been proposed in empty isometric space, namely, those embodying the experimental laws enunciated by Faraday and mathematically contained in the fundamental equations of Maxwell. If we define the second order electromagnetic force tensor F_{ij}, of anti-symmetric type, so that its six different components are the three electric and the three magnetic force components, X, Y, Z, L, M, N, and if σ denotes the density of electricity, these equations take the highly elegant tensor form of Maxwell-Lorentz:

$$\frac{\partial F_{ij}}{\partial \xi_k} + \frac{\partial F_{jk}}{\partial \xi_i} + \frac{\partial F_{ki}}{\partial \xi_j} = 0, \quad \frac{\partial F_i{}^a}{\partial \xi_a} = 4\,\pi\sigma u_i.$$

According to the laws found to hold, the electrical forces exerted on particles in motion are different than for particles at rest. The law is given by the tensor equation

$$f_i = F_{ia} u^a.$$

It seems highly probable that these equations are the simplest ones in agreement with the known facts.

Our second postulate for the detailed structure of matter is then merely that these equations do obtain. It may be stated as follows:

X. POSTULATE OF THE ELECTROMAGNETIC FIELD. *The ordinary Maxwell-Lorentz equations of the electromagnetic field hold.*

A more elaborate consideration, such as I have made elsewhere,[1] shows that the equations are in part deducible from more fundamental considerations; but, inasmuch as the phenomena of electricity and magnetism are not as much the subject of direct experience as are those of empty space and time, it appears to be best merely to state the bare hypothesis.

According to these equations, the energy density of electricity in isometric space-time is measured by the invariant

$$\frac{1}{8\,\pi}(X^2 + Y^2 + Z^2 + L^2 + M^2 + N^2).$$

---

[1] *Relativity and Modern Physics*, Chapter XII.

If this definition is adopted, the work done by the electromagnetic forces in moving matter will be equal to the loss of electromagnetic energy. Hence the principle of conservation of energy holds for electricity also.

At this stage we are led with Einstein to introduce an energy tensor of matter and electricity combined, namely,[1]

$$T^{ij} = \rho u^i u^j - p g^{ij} + \frac{1}{4\pi}\left(F^{i\alpha}F_{\alpha}{}^{j} + \tfrac{1}{4}F_{\alpha}{}^{\beta}F_{\beta}{}^{\alpha}g^{ij}\right)$$

as the general energy tensor for a perfect fluid. If we employ the same field equations as before, except that this modified tensor is used, it follows that the gravitational theory in the quasi-stationary state is the same as for the perfect fluid in the neutral state, *i.e.*, that electricity has no gravitational effect, while the fluid moves subject to the proper force due to pressure and the electrical force on the attached charge.

Thus the complete set of equations is

$$\begin{cases} R_{ij} - \frac{1}{2}Rg_{ij} = \chi T_{ij}, \\ \dfrac{\partial F_{ij}}{\partial \xi_k} + \dfrac{\partial F_{jk}}{\partial \xi_i} + \dfrac{\partial F_{ki}}{\partial \xi_j} = 0, \quad \dfrac{\partial F_{i}{}^{\alpha}}{\partial \xi_\alpha} = 4\pi\sigma u_i. \end{cases}$$

Here the undetermined function f is the only function at our disposal in further specifying the properties of matter. A particular choice will be made so as to satisfy the third and last postulate. Before introducing this postulate, it is desirable to make some important remarks concerning the nature of the above equations, and to simplify them as far as may be by the elimination of unnecessary field functions.

---

[1] Cf. *Relativity and Modern Physics*, p. 206.

In the first place, it is known to be possible to replace the tensor F_{ij} by a simpler electromagnetic four-potential ϕ_i, in such wise that the second set of equations is automatically fulfilled, by writing

$$F_{ij} = \frac{\partial \phi_i}{\partial \xi_j} - \frac{\partial \phi_j}{\partial \xi_i}.$$

The remaining equations take the form

$$R_{ij} - \tfrac{1}{2} R g_{ij} = \chi T_{ij}, \quad \frac{\partial F_i^{\ \alpha}}{\partial \xi_\alpha} = 4 \pi \sigma u_i.$$

in which it is understood that F_{ij} is replaced by its value in terms of ϕ_i.

But the second set of these equations permits us to eliminate σ and u_i, since from them we obtain

$$\sigma = g^{\lambda\mu} \frac{\partial F_\lambda^{\ \alpha}}{\partial \xi_\alpha} \frac{\partial F_\mu^{\ \beta}}{\partial \xi_\beta}, \quad u_i = \frac{1}{4 \pi \sigma} \frac{\partial F_i^{\ \alpha}}{\partial \xi_\alpha},$$

and conversely. Hence, if we understand by σ merely its expression in terms of the four-potential ϕ_i, and by p the value of $f(\rho)$, it appears that our complete equations take the unified form

$$R_{ij} - \tfrac{1}{2} R g_{ij} = \chi T_{ij}.$$

There remain only the 15 dependent variables

$$g_{ij};\ \phi_i;\ \rho.$$

There are essentially 10 equations to determine them. Now to the electromagnetic potential ϕ_i may be added any arbitrary gradient tensor $\partial S / \partial \xi_i$ without affecting the equations. Hence there are really only 14 dependent variables to be considered. Thus there will be an arbitrariness of 4 arbitrary functions in the general solution of the field equations, corresponding to the arbitrariness in the choice of a coordinate system.

In this simplified form the distinction between the equations in empty space and for the fluid disappear, since all of the functions involved are properly field functions.

Thus an important unification has been achieved. Furthermore, since a static solution of these equations corresponds to values of g_{ij}, ϕ_i, ρ, which do not depend on x_1, there will result values of u^i similarly independent of x_1. Hence in a static distribution there will in general exist steady electrical currents. Matter now appears as merely that part of space in which a certain field function ρ does not vanish.

A more careful examination of the equations of such a fluid shows that a certain quantity

$$\int \rho e^{-\int \frac{dp}{\rho}} dn$$

is conserved, no matter what external forces, such as those of the electromagnetic field, may be applied.[1] For example, in the case of free matter, the mass is conserved; for an incompressible fluid, the volume is conserved; but in general this quantity has no obvious physical significance.

On the other hand, it follows from the fundamental equations that electricity is a substance of invariable amount, and existing only as attached to the fluid.

---

[1] Cf. *Relativity and Modern Physics*, p. 100.

Thus we have two quantities analogous to density and they must stand in a relation of constant proportionality for all time at a particular particle of the fluid. Hence if we write

$$\rho e^{-\int\frac{dp}{\rho}} = \frac{\phi}{\sqrt{\pi}}\sigma$$

and thus define the function ϕ, it will be constant necessarily all along the world line of the particle. We shall call ϕ the *substance coefficient.* This name is justified by the fact that it is only defined where matter is, and is elsewhere entirely indeterminate. It cannot then be regarded as a dependent variable in our equations.

If we concede that the density of matter is always to be positive in sign, it is obvious that the substance coefficient is positive where the density of electricity σ is positive, and negative where σ is negative. We are now prepared to specify the form of the function f by means of the final special postulate.

XI. Postulate of Fluid Tension. *The tension* $-p$ *in the perfect fluid is proportional to the square of the electrical density* σ.

It is not obvious at the outset that it is possible to choose the function f so that this additional requirement is met, but we shall prove that it is possible to do so.

In such a fluid, the density and pressure of the fluid as well as the density of electricity may be constant when the fluid is at rest. Now no matter by what electromagnetic forces the fluid is brought into a new state of equilibrium, the substance coefficient, which was originally constant, must have the same constant value as it had at the outset. Accordingly the equation last written must hold if we replace p by $-\frac{1}{2}c^2\sigma^2$, according to the hypothesis of the postulate. Differentiating both sides of this equation logarithmically, there results the ordinary linear differential equation of the first order for ρ,

$$\frac{d\rho}{d\sigma} - \frac{\rho}{\sigma} = -c^2\sigma.$$

Integrating this equation, and recalling that if σ is 0, ρ will be also, we find

$$\rho = -c^2\sigma^2 + l\sigma.$$

Therefore we find the only possible relation between the pressure and density to be that given parametrically by equations of the type

$$p = -\tfrac{1}{2}k^2\tau^2, \quad \rho = \tau - k^2\tau^2,$$

where k is to be taken as a universal physical constant.

If we substitute these values of p and ρ in the equation defining the substance coefficient, it reduces to

$$\tau = \frac{\phi}{\sqrt{\pi}}\sigma.$$

Hence the tension is not only proportional to the square of the density of electricity, but also to the square of the substance coefficient, which seems reasonable enough.

In this way, and in this way only, the choice of the arbitrary function f entering in the constitutive equation can be made so as to satisfy the requirement of the last postulate.

At this stage it is important to notice several facts about such a perfect fluid. It is clear that the range of values of the parameter τ is from 0 to $1/k^2$. During this variation the tension constantly increases, while the density first increases to $1/4k^2$ and then decreases to 0 at the "critical tension" $1/2k^2$. Because of this diminishing density of the perfect fluid, there will be increasing responsiveness to electrical forces as the critical tension is approached. Since the total charge is invariable, the volume must diminish as the tension and electrical density increase. The situation is directly contrary to what is observed for the cohesive forces of ordinary fluids. It is not to be forgotten, however, that electricity is always present, so that there are counterbalancing repulsive forces. Evidently, since these repulsive forces tend to increase as the electrical density increases, the tensional forces operate so that stable equilibrium may be possible.

When the tension exceeds $1/4k^2$, $dp/d\rho$ is positive as is the case in ordinary fluids. The perfect fluid will then offer resistance similar to that of an ordinary fluid under pressure. It is this physical property which operates to maintain the separate identities of colliding portions of the fluid, whether they are oppositely or similarly charged.

The character of the appropriate boundary conditions must also be briefly referred to. At the boundary between a region occupied by the fluid and empty space it is natural to demand that the pressure and density of the fluid vanish. This is the case for ordinary fluids, and a more careful consideration based upon a study of thin transitional layers justifies this condition. Similarly, there must be continuous variation of pressure and density in passing from one part of the fluid to an adjacent part. Furthermore, the components of electric and magnetic force will be taken to vary continuously throughout inasmuch as the dielectric constant of the fluid is taken as unity.[1] The density σ of electricity, however, need not be continuous across the boundary between oppositely electrified parts of the fluid, although this density must vanish along a boundary in empty space.

With these considerations in mind, we are in a position to suggest a possible development of a theory of atomic structure.

Let us imagine a very large number of primordial units e of positive electricity attached to nuclei with a substance coefficient ϕ, constant throughout these nuclei. Under the action of mutual forces of repulsion, these will tend to scatter and thus to maintain their individual identities.

To begin with, let us investigate the possible condition of equilibrium of an isolated atomic nucleus of this sort. Suppose it may be taken to possess spherical symmetry. If r denotes the radial distance, which is the essential spatial variable, and if the very slight gravitational forces are ignored, the equations reduce to

$$\frac{dp}{dr} = \sigma F, \quad \frac{dF}{dr} + \frac{2\,F}{r} = 4\,\pi\sigma,$$

---

[1] Since the light wave has unit velocity in it.

where F stands for the radial electric force. The magnetic forces vanish of course.

If we express p in terms of σ, the first equation yields

$$-k^2\phi^2\frac{d\sigma}{dr} = \pi F.$$

Eliminating F from the second, we obtain

$$\frac{d^2\sigma}{dr^2} + \frac{2}{r}\frac{d\sigma}{dr} + \frac{4\pi^2}{k^2\phi^2}\sigma = 0.$$

If then we write $y = r\sigma$, we find

$$\frac{d^2y}{dr^2} + \frac{4\pi^2}{k^2\phi^2}y = 0,$$

a differential equation of familiar elementary type. The only solution of this equation which can yield an even function σ of r gives

$$\sigma = \frac{A}{r}\sin\frac{2\pi r}{k\phi}.$$

But σ must vanish at the boundary of the nucleus, which lies therefore at $r = nk\phi/2$ $(n = 1, 2, \cdots)$. Moreover, it is not possible that n exceeds 1, since σ must remain positive. Hence the radius of the positive nucleus is $k\phi/2$.

In order to ascertain the nuclear charge we have merely to evaluate the expression

$$4\pi\int_0^{\frac{k\phi}{2}}\sigma r^2 dr,$$

which can at once be integrated and yields the explicit formula

$$e = k^2\phi^2 A.$$

Likewise the mass of the nucleus is computed to be

$$M = \frac{\phi e}{\sqrt{\pi}} - \frac{e^2}{k\phi},$$

and the total static energy,

$$\mathcal{E} = 4\,\pi \int_0^\infty \left[\rho - p + \frac{F^2}{8\,\pi}\right] r^2 dr = \frac{\phi e}{\sqrt{\pi}} + \frac{e^2}{k\phi}.$$

If the variation of static energy (with given total charge e and substance coefficient ϕ) is obtained, it is found to vanish, and the ordinary tests for a maximum are fulfilled. Hence, if the static nucleus is disturbed radially, the static part of this energy is decreased. Since electromagnetic radiation of energy accompanies any such motion, the static energy must diminish towards some other minimum value. Therefore we conclude that this type of nucleus fails to fulfill the requirement of radial stability which is so essential.

Further consideration shows that this unstable type of nucleus will tend to contract in radius until its central part is filled with a sphere of the fluid, in the massless condition of minimum pressure and surrounded by a shell under pressure diminishing to zero at the outer boundary. It is this type of nucleus which seems to possess the required stability.

Now imagine that an electron with charge $-e$ and substance coefficient ψ approaches the nucleus under influence of the mutual electrical attraction. Experiment shows that ψ/ϕ is a very small quantity, — about $1/1840$, — so that the mass and radius of the electron are correspondingly smaller than for the nucleus. Of course, the electron is much more mobile because of its smaller mass, and so will develop a high velocity of approach, while that of the nucleus will not be affected very much.

The first conjecture which one naturally makes is that electromagnetic radiation will continue until the electron attaches itself to the nucleus in an outer spherical layer. But it is easily seen that this is impossible. In fact, at the outer edge of such a layer the density σ must vanish as before; and since the total charge is 0, the outside electrical force F will vanish also. Since F is continuous it is then necessary that $d\sigma/dr$ vanishes at the outer boundary. Hence it follows that σ and its gradient $d\sigma/dr$ both vanish there. This is only possible if σ vanishes identically, which cannot be.

Inasmuch as the electron has stability of the same sort as the nucleus, it appears that, in the final quiescent state after all possible dissipation of energy has taken place, both will take a position in mutual contact which is of much more stable type than either isolated from the other, since all of the electromagnetic energy of the external field, or nearly all of it, has been lost. If the electron is rotating about the nucleus under the influence of its attraction, the orbit will be nearly elliptic, and the electromagnetic effect may be nearly equivalent to that of an electric current in the same ellipse, so that radiation of energy takes place very slowly. It is even conceivable that under such rotation the electron must inevitably lengthen into a radiationless rotating anchor ring. Certainly the fluid tension in a ring of this kind would operate to hold it together.

Such a combination of a nucleus and an electron may be taken to be the hydrogen atom.

It has been pointed out previously that in any such theory as that here proposed, certain further boundary conditions will be necessary in order to supplement the differential equations, and so to prescribe somewhat the primordial distribution of matter and of electricity. Accordingly we shall assume that there exist also nuclei with substance coefficients $-\phi$ (the negative of that of the ordinary nuclei) and with charge $-e$. It will also be assumed that the number of positive nuclei is the same as the number of negative nuclei and electrons, while the number of positive nuclei is less than three times the number of negative nuclei by a considerable margin.

Hence, in the resultant statistical distribution under the influence of electrical attractions, it may be expected that the individual nuclear centers formed by the combination of positive and negative nuclei will carry a positive charge less than three times the negative charge; while outside of this nuclear center there will be a number of mobile attached electrons which neutralize the total nuclear charge. The notions of molecular combination on the basis of electropositive, electronegative, and neutral atoms, of a periodic classification of the various possible types of atoms or elements, are inherent in such an atomic structure, as an atomic theory like that of Lewis and Langmuir shows.

This relative frequency of positive and negative nuclei in the various chemical elements is demanded by the observed relationship between the atomic number of attached electrons and the atomic weights, namely, that the atomic number is at most half of the atomic weight. For if α is the number of positive nuclei and β of negative nuclei, then $\alpha - \beta$ is the atomic number, and $\alpha + \beta$ the atomic weight, so that

$$\alpha - \beta \leqq \tfrac{1}{2}(\alpha + \beta),$$

whence

$$\alpha \leqq 3\,\beta.$$

It is a striking fact that the approximate constancy of this ratio α/β appears as inevitable on the basis of the theory here advanced. In the first place a specific nuclear distribution of the electrically charged perfect fluid gives charges 1, 3, 5, . . . of opposite signs in successive layers of uniform thickness. This and other similar distributions indicate that the atomic weight $\alpha + \beta$ varies as the squared radius of the nuclear center. But the surface electrical force varies directly as the atomic number $\alpha - \beta$, and inversely as the square of the radius. Combining these results, it appears that the ratio $(\alpha - \beta)/(\alpha + \beta)$ measures the surface force of electrical attraction on a nuclear center. Hence if two centers approach in which the ratio is different, the one with larger electrical force will be likely to remove a negative nucleus from the other at collision, leaving behind an electron. Thus the new ratios will be more nearly equal. Hence gradually the ratios will approach approximate equality throughout, which requires that the ratio α/β is nearly 3 for all of the elements. If an incompressible fluid were employed as the carrier of electricity, such a constant ratio would not arise.

Radioactive disintegrations appear in this theory as evidence of incompleteness of statistical equilibrium between electrical forces at the surfaces of the nuclear centers.

The fact that the core of the center may be either a positive or negative nucleus indicates two values of the atomic number $\alpha - \beta$ (at most as large as $2\,\beta$) for each β. Thus the characteristic series of atomic numbers arises, with allowance for isotopes, but not requiring all atomic numbers to appear.

Let us carry the discussion further by inquiring somewhat into the atomic series, as it will be according to this theory.

A single positive nucleus carrying a charge which is a multiple of e will not come into existence, for computation indicates that such a multiple atom has a greater energy than its constituent parts.

This result shows that there will be at least two layers in the nuclear center, and it seems likely that there will be only two in the case next in order of simplicity.

Suppose then that there are p positive nuclei and q negative nuclei in the nuclear center. Let us suppose that the regions of the inner and outer layers are of thicknesses s and t. In each layer τ will satisfy the same differential equation as σ, since τ is a constant multiple of σ, and τ is continuous across the boundary of the two layers. An examination of the equation $\pi F = - k^2\phi^2\, d\sigma/dr$ shows too that $d\tau/dr$ has the same numerical value on the two sides (because F is continuous) but of opposite signs. Hence the graph of

τ is that schematically given by the figure below. Now if s has its maximum value $k\phi/2$, the same as the radius of the hydrogen nucleus, direct integration proves that the charge of the outer layer is three times that of the inner layer. If s diminishes towards its minimum value 0, then t approaches the radius of the hydrogen atom. In this limiting case the outer charge is infinitely greater than the inner. Thus it is clear that at any intermediate stage, like that indicated in the diagram, the outer charge will exceed three times the inner charge. Hence if we ask for the atom either of least atomic weight of this type, or of least total numerical charge, we conclude that its atomic weight is 4 with $p = 3$ and $q = 1$, and that there are two attached electrons.

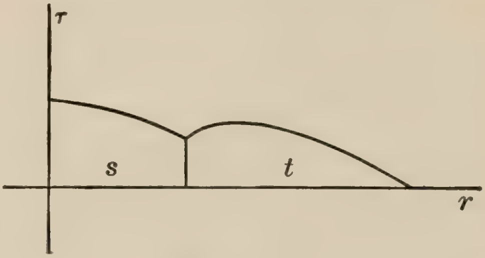

Fig. 21.

Such an atom, if contracted to stable form, would naturally correspond to the helium atom appearing next after hydrogen in the atomic series.

Moreover, the requirement that the positive nuclear charge is not more than three times the negative nuclear charge is fulfilled.

According to the principle of minimum energy, there is probably only one type of two-layer nucleus.

The next stage is clearly that of a three-layer nucleus, with either two positive layers and one negative, or two negative and one positive.

To go further in this direction would require more elaborate considerations. I propose to do so elsewhere. My principal purpose here has been to show how a very natural and simple series of assumptions of the purely mathematical kind might yield the kind of structure which gives the series of chemical elements and many of their wonderful properties. The specific theory formulated above is peculiar in that "negative nuclei" are introduced.

To sum up, then, we have endeavored to construct a mathematically complete theory of matter, according to which it is merely a perfect adiabatic fluid to which electricity is attached. The pressure of the particular type of fluid selected is so chosen that the pressure is negative (a tension) and is proportional to the square of the density of electricity; the electromagnetic forces are taken to obey the Maxwell-Lorentz laws; the space-time is assumed to be such as to yield the gravitational theory of Einstein. Everything else follows from the boundary conditions, which assume that the electricity is distributed in charges e and $-e$, so as to provide for positive nuclei, for negative nuclei of the same mass, and for negative electrons of an equal charge but of much lighter mass. These, aggregated in stable combinations, are taken to account for the various kinds of atoms of the chemical elements.

This is the kind of mathematical explanation of the constitution of matter toward which we must aim, if we believe with Pythagoras in the domination of nature by numerical law. But no theory of this type can hope to account for the phenomenon of life, for it will possess necessarily extremely narrow limitations of structural complexity.

No mention has been made of the quantum theory, despite its central interest for physicists to-day. This theory supplements the usual physical laws in order to explain spectroscopic phenomena. However, if a theory like the above were to lead to radiationless "stationary states," in which the electrons are deformed into closed rotating rings, then it might be found that such supplementary laws were consequences of the more familiar ones. As far as I know this possibility has not been taken sufficiently into account in the past. It should be thoroughly investigated.

CHAPTER VIII

THE PHILOSOPHICAL INFLUENCE OF RELATIVITY

It is the chief function of philosophy to furnish as wide and well-balanced a view as possible of nature and mind — those antipodal aspects of the universe between which reality lies. Advances in the physical and psychological sciences have always exerted a potent effect upon philosophic thought, and the reciprocal effect has been quite as noteworthy. For a mathematician like myself it is indeed a daring project to venture, by way of conclusion, into this speculative realm, and to attempt to evaluate the philosophical influence of the recent theory of nature called relativity.

By very definition, what is accidental is without especial interest for philosophy. From its earliest beginnings it has been directed towards that which has general significance in experience. Schools of philosophy fall into two classes: the first type takes its primary model in some generality of nature, and so may be called objectivistic; the second type begins with some generality of mind and may be called subjectivistic.

The penetration of nature with numerical relationships was doubtless noted very early by man. In truth, the most useful distinguishing mark of any class of objects is its number. Everywhere in nature, and also in mind, are classes and accordingly a role for number. Pythagoras and his school perceived that space and time were numerical in their essence. They conjectured too that number played a dominating part in other ways; the laws of the musical string and the motions of the stars and planets were sufficient indication to them that such was the case. The keynote of their various speculations was that the universe was arithmetical throughout.

Plato was the first great subjectivistic philosopher. He took as primary the model of abstract mathematical truth. His conception of the occupation of the Deity was that of continual geometrizing. It was his great advance to have appreciated fully that there were other ideas, not reducible to number or form, and yet having the same degree of living reality; chief among them were such social ideas as the good and the beautiful, which had been so highly stressed in the Socratic teachings. Material fact was held by him to be less real than these ideas themselves.

His successor, Aristotle, was inclined to a more objectivistic view, since he wished to ascertain the fundamental generalities by direct observation as well as by pure reasoning. He was led to the invention of formal logic, no doubt with the exemplar of geometry in mind, and defined the fields of the various sciences.

Thus Greek thought advanced far enough to appreciate the penetration of nature and mind with law. Perhaps its incompleteness was most conspicuous on the subjective side.

With the Renaissance, philosophers began to examine more closely the immediate deliverances of inner experience. The familiar dictum of Descartes, *Cogito ergo sum*, illustrates the new type of starting point. This introspective attitude created an immediate extraordinary change of appearances. Descartes proposed to account for the difference by adopting a dualistic view of the universe in which nature and mind were opposed. Leibniz envisaged instead a world of self-subsistent thought-monads. Both of these men appear to have been somewhat misled by their mathematical successes, so that they overestimated the power of pure reasoning. Contrary to the Aristotelian spirit, they attempted an *a priori* treatment of the universe which seems to-day largely formal arrangement.

One difficulty which philosophy was then laboring under deserves to be explicitly pointed out, namely, the uncritically accepted notion that objects were seen even as they were in essence. For this reason, Locke achieved a substantial advance by classifying certain properties of bodies, like size and shape, as primary, and other qualities, like color, as secondary since they exist only as sensation. Berkeley's doctrine that *esse* is *percipi* went still further in the same direction, while Hume adopted the extreme position that nought was real save impressions of sense and ideas.

The psychological ideas of Locke and the physical ideas of Newton, in particular the latter's corpuscular theory of light, fixed an apparently impassable gulf between the knower and the known. Nevertheless, space and time were still held to have an *a priori* character of necessity. It was in this way that Kant, near the beginning of the eighteenth century, was led to his objectivistic phenomenalism with the knower forever separated from the thing-in-itself, and perceiving it only under space and time as a form of the intuition.

Hegel abandoned the Kantian idea of an inaccessible reality behind knowledge. He centered on the dynamic aspect of consciousness in its continual transformation of the immediate sensory data. More recently, Bergson has analyzed consciousness from a similar position. For him also the starting point and goal lie in directly given knowledge, but he finds the essence to lie in a creative evolution of thought in duration, this duration being unmodified by treatment as an abstract moment of time. Thus Hegel and Bergson achieve a return to complete Berkeleian subjectivity.

It is apparent that, from Descartes to Bergson, all the philosophers mentioned have taken their point of departure in a new psychological insight.

Similarly, the development of biological thought has been the inspiration of the philosophy of Spencer and of a whole correlated movement.

Thus it appears that the main lines of approach to nature and mind are mathematical, physical, biological, psychological, and social, and constitute a kind of hierarchy. In each, a few characteristic notions are fundamental, while all else is relegated to a minor place. The following classification (reminiscent of that given by Comte) indicates each of these main divisions, a corresponding typical system of philosophy and some of the leading ideas:

Mathematical, Absolute Realism;
  Class, Relation, Inference, Abstraction.
Physical, Materialism;
  Space-Time, Matter, Electricity, Uniformity.
Biological, Detailed Naturalism;
  Organism, Stimulus, Function, Evolution.
Psychological, Positivism;
  Sensation, Memory, Will, Idea.
Social, Ethical Idealism;
  Personality, Freedom, Value, Ideal.

Of these, the earlier are objective, the later, subjective in outlook. The test of the objectivity of the mathematical and physical levels in this nature-mind spectrum lies in the fact that active thought at these levels goes on without any explicit reference to mind; the thoroughgoing subjectivity of the psychological and social levels is revealed by the opposite condition; the biological level is clearly of an intermediate type.

The position of a technically developed philosophical system may fall at an intermediate level; for example, pragmatism is biological-psychological in its structure. In all philosophies, supreme "reality" is attached to the primary level, while the secondary levels are held in some sense to be derivative. For instance, on the biological level, inference is merely a process of thought which "works"; and, on the psychological level, space appears as a complex of visual and tactual associations.

The naming of the nature-mind spectrum is a matter of interest. If we are objectivistic in tendency we call it the Absolute; if subjectivistic, Knowledge.

In my opinion, the fact that these levels are independent to the extent of being largely self-sufficient indicates that the best method for the philosophic grasp of the whole is to take the various levels as of coordinate importance and reality. At least this is the reasonable philosophic view of most value for the ordinary man, even if it be merely a kind of naïve realism.

The greatest direct philosophical influence of the physical theory called relativity lies at the physical level in the nature-mind spectrum, for it affects the meaning of the fundamental terms.

The classical concepts of absolute space and absolute time are replaced by a combined space-time. Time apart from space and space apart from time lose their intrinsic significance. The new element is the event, and the new underlying basis is the four-dimensional space-time continuum of events.

The matter and electricity in space-time are now thought of as affecting its very structure. Uniformity of physical law still holds, but this uniformity requires effects in the physical realm to be locally transmitted in the spatio-temporal sense, whereas, in the older view, the " state " of the physical universe at one " instant " affected the immediately subsequent states throughout all space.

It was mainly to the careful consideration of these influences of relativity upon the philosophy of nature that the earlier chapters were devoted.

Whitehead [1] has carried the new view of nature back to the mathematical (*i.e.*, logical) level, by showing how space and time arise from his standpoint. In consciousness there is the directly " discerned " and also the ultimately " discernible." Classes of events are what are discerned, and these are interrelated through compresence or lack of compresence in consciousness. Thus an event is a " slab of nature." The technical relation between events is that of spatio-temporal inclusion. The method of extensive abstraction then yields the notion of the idealized event or point-instant as a collection of the slab-like events mutually in the relation of inclusion. The method is such that the space and time determined are clearly relative to the means of measurement employed. A point is merely a collection of events with the same three space numbers; an instant is a collection of events with the same time number.

---

[1] *The Concept of Nature*, Cambridge, England, 1920.

As is indicated by Whitehead, the approach to space-time given in these general terms becomes definitely mathematical under technical elaboration. In this form it is an abstraction of space-time, capable of giving the same kind of service for the new theories as various other logical systems perform for ordinary space; it is similar in structure to Huntington's system for ordinary space, in which the sphere is the undefined element, while the undefined relation is inclusion. It is on this account that the analysis belongs entirely at the mathematical level.

With space-time so analyzed, the object can be defined as a collection of events united by a certain kind of common modality. The aim of the physical sciences is then to specify the uniformities underlying the behavior of such objects in space-time, in terms which are essentially independent of the arbitrary method of space-time measurement selected.

It is plain that Whitehead builds in this way a mathematical approach to the new theories, of philosophical type just because the general logical aspects, rather than the technical mathematical ones, are stressed.

Similarly, Bergson [1] has pointed out an interesting parallelism between his own speculations and the later theories of relativity. He had taken the notion of duration, unmodified by intellectual analysis, to lie at the basis of experience. Now in the new theories the notion of the measurable interval of local time serves as the basis for the subsequent technical construction of measurable space and time. Hence relativity affords a certain independent justification of the validity of Bergson's prior conclusions. Moreover he has called attention anew to the reality of the notion of duration from the psychological point of view, in that two beings can identify themselves in thought, and so in duration. Yet there are two kinds of simultaneity, as in relativity. "The first," says Bergson, "is interior to events, it makes part of their materiality, it originates in them. The other is simply attached to them by an observer exterior to the system. The first expresses something about the system itself; it is absolute. The second is changing, relative, fictitious, it originates in distance."

For the relativist the flow of time for all beings is such that it is not possible to conceive of this multiple time as a simple flow without distorting its inner nature. To a certain extent, therefore, Bergson's refusal to treat time abstractly acquires justification.

---

[1] *Durée et Simultanéité*, Paris, 1922.

These reactions of Whitehead and Bergson are sufficient to indicate that relativity is proving very suggestive at the mathematical and psychological levels.

Haldane[1] has attempted to apply the "principle of relativity," even at the social level, in dealing with society and the state. But here he is referring to a long-established principle largely to be found in Plato, as the following passage of Haldane's makes manifest:

> The real lesson which the principle of the relativity of knowledge teaches us is always to remember that there are different orders in which both our knowledge and the reality it seeks have differing forms. These orders we must be careful to distinguish and not to confuse. We must keep ourselves aware that truth in terms of one order may not necessarily be a sufficient guide in the search for truth in another. We have, in other words, to be critical of our categories.

This is substantially only the doctrines of the differing levels of thought, stated above.

There have been many attempts to read into the Einstein theories a significance which does not belong to them.

It has frequently been thought that the use of an "observer" in his development of the special theory of 1905 indicated the presence of a highly subjectivistic element in physics itself, thus reinforcing the subjectivistic point of view in philosophy. But it is very easy to phrase the new theories so as to eliminate the "observer" entirely. For instance, instead of stating that "an observer A signals to B by a light flash, whereupon B immediately signals back to A," one may say that a light flash goes from a source at A to a reflecting surface at B and so back to A. By such a modification of Einstein's exposition the use of an observer is easily and naturally avoided, as much as it can be in any physical theory.

---

[1] *The Reign of Relativity*, New Haven, 1921.

There is another claim too, made by Einstein himself, that the laws of nature must be invariant in form under all transformations of coordinates. This is a confusing form of statement. The fact is that, while the language of tensor analysis may be convenient for use in the formulation of the laws of nature, the possibility of such formulation has no bearing on the character of physical laws. For example, those of classical physics can also be completely expounded in the language of tensors. The following simple analogy will serve to show how this can be possible. A plane is a very specialized type of surface, and yet it can be completely characterized by the use of general ideas applicable to all curved surfaces. Note first that the angle between two intersecting lines on a curved surface has a meaning; note further that the length of a line on such a surface has a meaning, and that there is a shortest line between any two points — the geodesic — which is the straight line in the plane. Hence the plane can be characterized as the surface such that the three geodesics joining any three points intersect in three angles whose sum is precisely a straight angle. Clearly this characterization is couched in general terms, and yet is complete as far as the interior properties of the plane are concerned.

Also, despite the unusual generality of the formulation of laws in the theory of relativity, its physical content is as specific as that of previous theories. For example, the rotation of a small body has a perfectly definite significance, although the opinion has been widespread, in accordance with the claim of Einstein himself, that rotation is " relative." It is doubtless very reasonable that if there were only a single central body it could not be rotating — for with respect to what could it be rotating? Nevertheless the theory declares that it may be rotating, namely, with respect to beams of light sent out from the body. Hence, as it stands, the general theory of relativity fails to meet a plausible philosophical requirement.

Yet it is specific in all respects, even if formulated in a language of greater generality than was useful in the old theory. Now it may be that, when the laws of physics are formulated in terms of arbitrary coördinates, they appear as the simplest laws qualitatively consistent with the known facts. If this be the case, a mystical truth of high importance is revealed; and this is perhaps the truth which Einstein had in mind.

There is, however, a sense in which relativity can be made to throw further light upon the mathematical-logical factor in mind and nature, namely, as a typical instance of the abstraction. This is only the case because the familiar classical abstractions of space and time have been held sacrosanct and qualitatively different from all other abstractions. But the theory of relativity has made it clear that it is not yet possible to decide what will be the final abstract form which physical theories will take, even in dealing with space and time.

In consequence, certain elementary truths about abstract structures at any level of thought have not been properly appreciated in the past. It is to these truths, as instanced in relativity and applicable in a measure to philosophy, that I propose to turn now.

The word " abstraction " is to be interpreted here in a very wide sense. When a primitive man looked at the sun and the moon, it must have been obvious in his mind that there was a similarity about the two objects. Here the perception of likeness amid difference is the nascent intuition, the forerunner of the symbolic abstraction. The same fact in the language of the early Greeks would have been expressed by saying that the sun and moon were spherical in shape. But the entire meaning of such a statement was latent in the perception of the primitive man. Moreover, even the Greeks used various intuitions in their geometric reasoning which were not explicitly stated. Within recent years, however, symbolic logic has become available as a calculus (*i.e.*, a very brief complete language in which the words are symbols) with which to deal with abstractions. Certain postulates (*i.e.*, intuitive inferences) are stated in explicit form, and then by a chain of inferences all the latent meaning of the systematized abstraction is developed. Thus the abstraction appears in three stages: the nascent or purely intuitive stage, the stage of systematic elaboration and application, and the completely symbolic stage.

The various stages of an abstraction can be given a very simple exemplification in the abstraction of local time. The first stage is that in which it is perceived that the order of events will be *A*, *B*, *C*, etc. In the second stage the notions of "before" and "after" will have been embodied in the symbolic word so that *A* will be declared to happen before *B*, etc.; at the same stage it will be held intuitively obvious that if *A* happens before *B*, and *B* happens before *C*, then *A* happens before *C* — a postulate; events will be numbered, so that arithmetized time is available in dealing with events. In the final symbolic stage the notion of an instant of time will be arrived at by abstraction of the time continuum. It will be proved that with the class of "instants" as undefined elements, and with "before" as undefined relation, and obvious postulates, the time continuum can be completely characterized. This is the third and last step. The final abstraction seems to stand out and exist at the mathematical level, apart from any concrete embodiment of it, in a very remarkable way.

Thus, properly interpreted, the abstraction is of transcendent importance. Only by means of the complementary processes of abstraction and experiment is it possible to extend knowledge.

We are thus led to state as a first principle:

I. Abstractions originate in the domain of the intuitively perceived.

The second general principle is essentially that given by Haldane, and may be stated as follows:

II. Every abstraction is to be applied in its appropriate domain of validity.

For example, the geometry of Euclid idealizes static spatial relations although none are exactly static. The special relativity holds only for empty space-time, and, while space-time is nearly empty of matter in certain parts, it is not exactly so. Likewise, the general relativity assumes that there is no limit to the indefinitely fine structure of matter.

However, all these particular abstractions apply at the physical level. At the other levels the abstractions used are far less exactly applicable, in particular because the meaning of the various terms is not susceptible of equally precise technical definition.

This seems to be more or less in the nature of the case, since only physical entities are presumed to be elementary, while those of biological, psychological, or social type are statistical (although not less important on that account).

III. As more complete abstractions are made, they may be expected to include their predecessors.

Thus, in a sense specified previously, the special theory of relativity includes the classical theory, as the case of very small relative material velocities; and the general theory of relativity includes the special, as the case when matter is very sparsely distributed.

In general, the advances in the physical and psychological sciences, as well as in philosophy, suggest the

Hegelian principle that new abstractions contain their predecessors in some sense or other.

IV. The undefined elements, relations, and postulates of a particular abstraction are to a large extent arbitrary.

For example, the solid may be made the element in ordinary geometry, as well as the point. In the latter case the undefined relation may be " betweenness " or " separation." Similarly, in relativity we may take the four-dimensional extended event or the point-instant as the element; and in the latter case the undefined relation may be " local time " or " local distance." The range of choice is always very large, and there are many interesting possibilities besides those referred to.

Hence, even though the undefined elements, relations, and postulates are selected so as to have a large degree of self-evidence, the choice made is by no means determined by this requirement. If two available sets of undefined elements, relations, and postulates are considered, either can be completely interpreted in terms of the other. Hence actual difference in conclusions does not follow from the difference of terminology. The essential item is that all elements, relations, and postulates from the one system can be systematically interpreted in the other, and *vice versa*.

On this account, the fact that there are many philosophical systems does not necessarily imply any essential differences between them, except in so far as the meaning of the terms used cannot be consistently interpreted back and forth. For example, suppose that one philosophical system regards as real only what lies at the physical level, and a second such system holds to the superior reality of the psychological level. Notwithstanding this difference, it may be possible for either to interpret technically the kind of reality used by the other. The mere difference in usage of words is not significant. Of course, if the abstractions have reference to ascertainable fact and differ in a substantial way, experiment will reveal which of the two is correct. However, since systems of philosophy do not differ about the facts at the various levels as much as about their interpretation, it seems patent that such systems may be abstractions with different undefined elements, relations, and postulates, but otherwise in satisfactory harmony.

V. The usefulness of an abstraction is relative to its inherent simplicity of structure and its agreement with the facts.

For example, the usefulness of the theory of relativity depends on the circumstance that it possesses the same inherent simplicity as the classical theory, while it explains more facts than that theory did. Similarly, a philosophical system possesses merit only in proportion to its simplicity and capacity of accounting for the facts of mind and nature.

The five general principles above ought in particular to apply to the social abstractions, which deal with ethical and religious questions, and have a large importance for the individual and for the race. Like the abstractions of space and time, some of them have often been held final and unmodifiable. It is accordingly very natural to conjecture that this opinion is not justified. Let us attempt to apply those principles at the social level.

Inasmuch as all abstractions develop from an intuitive phase, genuine intuitions in ethical or religious matters may safely be regarded as the forerunners of subsequent abstractions. They must therefore be given weight and consideration, and yet must be examined with care. The existence of many social intuitions is unquestionable.

The limitation of the domain of validity of a particular abstraction and its tendency to be included in a larger one, suggest the gradual evolving of the accepted social code in such wise that all values are retained during the process. Relativity does not suggest then that ideals are relative and shifting, but rather that they will enlarge from time to time.

The great variety of ways in which one and the same abstraction may be approached indicates that the real differences of underlying opinion at the social level can only be inferred when there exist practical decisions that are different. Undoubtedly this fact should make for sympathy and for tolerance. A similar conclusion is sometimes drawn from the supposedly subjectivistic character of relativity.

Finally, social abstractions ought to be simple and straightforward, designed so as to be actually applicable under wide conditions.

Thus the true role of the systematized abstraction has become more apparent, and freedom has been won to use a formula or a theory and yet not be enslaved to it. Undoubtedly relativity has acted as a salutary influence in this direction upon the entire field of human thought.

INDEX

A

Abstractions, 176–182
Ahmes, 2
Aristotle, 7, 8, 14–15, 44, 167
Astronomy, 8–11, 41
Atom, 17, 139, 160, 164
Atomic number, 160–161
Atomic weight, 160–161

B

Berkeley, 168
Bergson, 169–174
Bodies, 5–8, 27–28, 29–30, 39–40, 69–70

C

Campbell, W. W., 87
Causal cone, 117
Causal equation, 116
Causation, 114–118
Clocks, 37–38, 43–44, 75–77, 80–81, 97–99, 101
Comte, 168
Coordinates, 111, 175–176
Copernicus, 10
Curvature, 63, 65–66, 131–132

D

Descartes, 11, 51, 168, 169
Distance, 32, 44, 112–113
Doppler, 78
Dynamics, 135–138

E

Eddington, 68, 88
Einstein, 1, 21–23, 41, 47, 50, 52, 55, 58, 64, 68, 70, 71, 72, 83–88, 105, 106, 129, 132, 150, 174–176
Eisenhart, 68
Electricity, 147–149, 152–153
Electromagnetism, 15–16, 57, 59, 79, 83–84, 148–150
Electron, 139, 158–160
Energy, 105, 136, 138, 147, 149
Energy tensor, 146, 150
Equivalence, principle of, 129–131
Ether, 16
Ether drift, 17–18, 81–83
Euclid, 1, 5–8, 10, 16, 23, 68, 69, 179
Events, 42, 62–63, 73, 75, 105, 136, 138

F

Faraday, 1, 15–16, 57, 77, 78, 148
Field equations, 132, 143, 150–152
Field functions, 112–114, 151–152
Fitzgerald, 19
Fizeau, 81
Fluid, 143–146, 153–156
Force, 10, 29, 50, 51, 52–56, 144–145, 149

G

Gale, 82
Geometry, 2–5, 31–37, 47–48
Gravitation, 10–11, 47, 50–52, 55–57, 59–62, 66–68, 131–133

H

Haldane, 174, 179
Hegel, 169
Helium, 160–162
Hipparchus, 9
Hume, 168
Huntington, 172
Hydrogen, 158–160

I

Isotopes, 162

K

Kakeya, 4
Kant, 168
Kepler, 10, 50

L

Langmuir, 160
Leibniz, 51, 168
Lewis, 160
Light, 9, 16, 28, 43, 79
Light, age of, 101, 102
Light, bending of, 86, 132–133
Light, velocity of, 9, 16, 26, 28, 91
Locke, 168
Logic, 167, 177–178
Lorentz, 19–21, 83

M

Mass, 50, 104–105
Matter, 66, 152
Matter, free, 60, 140–143
Matter, structure of, 30, 139, 160, 164
Maxwell, 15–16, 18, 57, 79, 83, 148–150
Maxwell-Lorentz equations, 148–150
Michelson, 17, 82
Michelson-Morley experiment, 17–18, 20, 79–82, 90
Miller, D. C., 81–82
Minimum principle, 136–137
Minkowski, 62, 84
Morley, 17

N

Newton, 1, 11–15, 23, 40, 51, 52, 56, 59, 68, 69, 77, 78, 85, 86, 99
Nucleus, negative, 156–158
Nucleus, positive, 160
Number, 1–2, 166–167

P

Particles, 42–43, 119, 120–128
Perihelion of Mercury, 85, 132–133
Philosophy, history of, 166–169
Philosophy, types of, 166, 169–171, 180–181
Plato, 167
Poincaré, 19, 136, 144
Postulates of physics, 110, 112, 114, 118, 128, 130, 133, 145, 149, 153
Potential, electromagnetic, 151–152
Potential, gravitational, 60
Ptolemy, 9
Pythagoras, 4–5, 9, 164, 167

Q

Quantum theory, 165

R

Relativity, 1, 21–23, 24, 45–49, 77–78, 89–99, 100–101, 171
Relativity, geometric, 8
Relativity, influence of, 166, 171–182
Relativity, mechanical, 14
Relativity, optical, 21–22
Relativity, paradoxes of, 99–107
Relativity, principle of, 52–58, 67, 83–84, 97, 106, 174
Römer, 26
Rotation, 12, 14, 176
Russell, H. N., 87

S

St. John, 87–88
Simultaneity, 22–24, 26–27, 38–42, 91, 173
Space, 16, 22–23, 43
Space, absolute, 6–8, 11, 13, 16–18, 41, 42
Space, curved, 105–107
Space, earth-, 28, 42
Space, finite, 106–107
Space, relative, 6–8, 11, 30–31
Space-time, 23, 25, 42, 43, 44, 46, 62–63, 72–75, 78, 95, 97, 110–112, 171–173
Space-time, isometric, 118–128
Spectral shift, 87–88, 132–133
Stability, 133–138
Substance coefficient, 152–153

T

Tensors, 62, 130, 140–141
Thales, 4
Time, 1, 21–24, 25–26, 37–38, 40–42, 173–174
Time, absolute, 9, 11, 18–19, 41–42
Time, earth-, 27, 28, 42
Time, local, 21, 84–85, 88, 128–129, 178
Time, relative, 11, 37–38

V

Veblen, 69

W

Weyl, 69
Whitehead, 171–173, 174
World line, 63

10-30-81

SCIENCE AND ENGINEERING LIBRARY
University of California, San Diego

Please Note: This item is subject to
RECALL after one week.

DATE DUE

OCT 7 1982
SEP 28 REC'D
NOV 6 1982
JAN 16 REC'D
DEC 23 1985
DEC 16 1985 REC'D
DEC 16 REC'D

SE 7 (Rev. 7/82)
UCSD Libr.